Priebe
09/902432
Seq. IDs 3 & 4 w/ Interf

```
> O <
O| |O  IntelliGenetics
> O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file us-09-902-432-3.res made by bshears on Fri 13 Dec 102 15:28:18-PST.


Query sequence being compared:US-09-902-432-3 (1-6160)
Number of sequences searched:                    2
Number of scores above cutoff:                   2

      Results of the initial comparison of US-09-902-432-3 (1-6160) with:
        File : /home/bshears/*97827*seq

   100-
      -
N     -
U   50-
M     -
B     -
E     -
R     -
      -
O     -
F   10-
      -
S     -
E    5-
Q     -
U     -
E     -
N     -
C     -
E     -                            *                                *
S    0-----------------------------------------------------------------------
      |     |     | |     |     |      ||     |     |     |     | |
SCORE 0    578   1156   1733   2311   2889| 3467   4044   4622   5200
STDEV              -1                     0                      1

                              PARAMETERS

Similarity matrix       Unitary    K-tuple                  4
Mismatch penalty              1    Joining penalty         30
Gap penalty                1.00    Window size             32
Gap size penalty           0.33
Cutoff score                  0
Randomization group           0

                           SEARCH STATISTICS

Scores:                  Mean       Median     Standard Deviation
                         3825        2452      1943.84

Times:                    CPU                  Total Elapsed
                      00:00:00.02              00:00:00.00

Number of residues:                10274
Number of sequences searched:          2
Number of scores above cutoff:         2


The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.


The list of best scores is:
```

```
                                                            Init. Opt.
Sequence Name     Description                        Length Score Score  Sig. Frame
-----------------------------------------------------------------------------------
   1. US-08-978-277A-3 Sequence 3, Application U    5200  5200  5200   0.71   0
   2. US-08-978-277A-1 Sequence 1, Application U    5074  2451  4579  -0.71   0


1. US-09-902-432-3 (1-6160)
   US-08-978-277A-3 Sequence 3, Application US/08978277A

 Sequence 3, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  3:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  5200 base pairs
      TYPE:  nucleic acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  cDNA
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:
    ORIGINAL SOURCE:

Initial Score    =   5200  Optimized Score  =   5200  Significance =  0.71
Residue Identity =   100%  Matches          =   5200  Mismatches   =     0
Gaps             =      0  Conservative Substitutions              =     0

            10        20        30 X      40        50        60        70
  CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGC
                                  ||||||||||||||||||||||||||||||||||||||||
                                  ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGC
                                  X        10        20        30        40

          80        90       100       110       120       130       140
  AGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAG
            50        60        70        80        90       100       110

       150       160       170       180       190       200       210
  CAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
      CAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCA
          120       130       140       150       160       170       180

      220       230       240       250       260       270       280
      ACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      ACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTG
          190       200       210       220       230       240       250

  290       300       310       320       330       340       350       360
      ATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      ATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT
      260       270       280       290       300       310       320

          370       380       390       400       410       420       430
      CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTT
  330       340       350       360       370       380       390       400

        440       450       460       470       480       490       500
      CAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTAT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      CAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTAT
          410       420       430       440       450       460       470

        510       520       530       540       550       560       570
      TTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      TTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCA
          480       490       500       510       520       530       540

        580       590       600       610       620       630       640
      CTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      CTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGG
          550       560       570       580       590       600       610

  650       660       670       680       690       700       710       720
      AGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      AGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC
      620       630       640       650       660       670       680

          730       740       750       760       770       780       790
      TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCA
  690       700       710       720       730       740       750       760

          800       810       820       830       840       850       860
      AAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      AAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACA
          770       780       790       800       810       820       830

        870       880       890       900       910       920       930
      GTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      GTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCA
          840       850       860       870       880       890       900

      940       950       960       970       980       990      1000
      AGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      AGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGG
          910       920       930       940       950       960       970

 1010      1020      1030      1040      1050      1060      1070      1080
      AAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      AAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT

       980       990      1000      1010      1020      1030      1040

          1090      1100      1110      1120      1130      1140      1150
      CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGA
 1050      1060      1070      1080      1090      1100      1110      1120

        1160      1170      1180      1190      1200      1210      1220
      AGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      AGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCC
          1130      1140      1150      1160      1170      1180      1190

      1230      1240      1250      1260      1270      1280      1290
      ACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      ACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAG
        1200      1210      1220      1230      1240      1250      1260

      1300      1310      1320      1330      1340      1350      1360
      AAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTG
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      AAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTG
        1270      1280      1290      1300      1310      1320      1330

 1370      1380      1390      1400      1410      1420      1430      1440
      CTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      CTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC
      1340      1350      1360      1370      1380      1390      1400

          1450      1460      1470      1480      1490      1500      1510
      CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAA
 1410      1420      1430      1440      1450      1460      1470      1480

        1520      1530      1540      1550      1560      1570      1580
      GAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      GAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGA
          1490      1500      1510      1520      1530      1540      1550

      1590      1600      1610      1620      1630      1640      1650
      AGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      AGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCC
          1560      1570      1580      1590      1600      1610      1620

      1660      1670      1680      1690      1700      1710      1720
      CAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTC
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      CAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTC
        1630      1640      1650      1660      1670      1680      1690

 1730      1740      1750      1760      1770      1780      1790      1800
      TGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      TGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA
      1700      1710      1720      1730      1740      1750      1760

          1810      1820      1830      1840      1850      1860      1870
      AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTT
 1770      1780      1790      1800      1810      1820      1830      1840

        1880      1890      1900      1910      1920      1930      1940
      CTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGT
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
      CTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGT
          1850      1860      1870      1880      1890      1900      1910
```

```
   1950      1960      1970      1980      1990      2000      2010
 CAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATA
     1920      1930      1940      1950      1960      1970      1980

 2020      2030      2040      2050      2060      2070      2080
 CTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAG
   1990      2000      2010      2020      2030      2040      2050

2090      2100      2110      2120      2130      2140      2150      2160
 ATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG
 2060      2070      2080      2090      2100      2110      2120

     2170      2180      2190      2200      2210      2220      2230
 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAG
2130      2140      2150      2160      2170      2180      2190      2200

    2240      2250      2260      2270      2280      2290      2300
 CGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAA
     2210      2220      2230      2240      2250      2260      2270

   2310      2320      2330      2340      2350      2360      2370
 AATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGA
    2280      2290      2300      2310      2320      2330      2340

 2380      2390      2400      2410      2420      2430      2440
 GTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGC
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 GTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGC
   2350      2360      2370      2380      2390      2400      2410

2450      2460      2470      2480      2490      2500      2510      2520
 AAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG
 2420      2430      2440      2450      2460      2470      2480

     2530      2540      2550      2560      2570      2580      2590
 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGC
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGC
2490      2500      2510      2520      2530      2540      2550      2560

    2600      2610      2620      2630      2640      2650      2660
 TGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTG
     2570      2580      2590      2600      2610      2620      2630

   2670      2680      2690      2700      2710      2720      2730
 ATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTC
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 ATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTC
    2640      2650      2660      2670      2680      2690      2700

 2740      2750      2760      2770      2780      2790      2800
 TTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAA
   2710      2720      2730      2740      2750      2760      2770

2810      2820      2830      2840      2850      2860      2870      2880
 CTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT
 2780      2790      2800      2810      2820      2830      2840

     2890      2900      2910      2920      2930      2940      2950
 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGG
2850      2860      2870      2880      2890      2900      2910      2920

     2960      2970      2980      2990      3000      3010      3020
 GGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 GGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAG
     2930      2940      2950      2960      2970      2980      2990

   3030      3040      3050      3060      3070      3080      3090
 CCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCC
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCC
     3000      3010      3020      3030      3040      3050      3060

 3100      3110      3120      3130      3140      3150      3160
 TCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGT
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGT
   3070      3080      3090      3100      3110      3120      3130

3170      3180      3190      3200      3210      3220      3230      3240
 CAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG
 3140      3150      3160      3170      3180      3190      3200

     3250      3260      3270      3280      3290      3300      3310
 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTA
3210      3220      3230      3240      3250      3260      3270      3280

    3320      3330      3340      3350      3360      3370      3380
 CTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGC
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGC
     3290      3300      3310      3320      3330      3340      3350

   3390      3400      3410      3420      3430      3440      3450
 TCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAA
    3360      3370      3380      3390      3400      3410      3420

 3460      3470      3480      3490      3500      3510      3520
 GCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 GCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTA
   3430      3440      3450      3460      3470      3480      3490

3530      3540      3550      3560      3570      3580      3590      3600
 AAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT
 3500      3510      3520      3530      3540      3550      3560

     3610      3620      3630      3640      3650      3660      3670
 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTA
3570      3580      3590      3600      3610      3620      3630      3640

     3680      3690      3700      3710      3720      3730      3740
```

```
CACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTG
      3650      3660      3670      3680      3690      3700      3710

  3750      3760      3770      3780      3790      3800      3810
ACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGG
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
ACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGG
    3720      3730      3740      3750      3760      3770      3780

3820      3830      3840      3850      3860      3870      3880
CTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGC
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
CTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGC
  3790      3800      3810      3820      3830      3840      3850

3890      3900      3910      3920      3930      3940      3950      3960
 AGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC
 3860      3870      3880      3890      3900      3910      3920

       3970      3980      3990      4000      4010      4020      4030
 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGA
3930      3940      3950      3960      3970      3980      3990      4000

     4040      4050      4060      4070      4080      4090      4100
 CACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACC
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACC
      4010      4020      4030      4040      4050      4060      4070

  4110      4120      4130      4140      4150      4160      4170
 AAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTG
    4080      4090      4100      4110      4120      4130      4140

 4180      4190      4200      4210      4220      4230      4240
 TGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTAT
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTAT
   4150      4160      4170      4180      4190      4200      4210

4250      4260      4270      4280      4290      4300      4310      4320
 TACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG
 4220      4230      4240      4250      4260      4270      4280

       4330      4340      4350      4360      4370      4380      4390
 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACC
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACC
4290      4300      4310      4320      4330      4340      4350      4360

     4400      4410      4420      4430      4440      4450      4460
 TACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATG
       4370      4380      4390      4400      4410      4420      4430

   4470      4480      4490      4500      4510      4520      4530
 CTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTA
     4440      4450      4460      4470      4480      4490      4500

 4540      4550      4560      4570      4580      4590      4600
 AGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCG
   4510      4520      4530      4540      4550      4560      4570

4610      4620      4630      4640      4650      4660      4670      4680
 AAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT
 4580      4590      4600      4610      4620      4630      4640

       4690      4700      4710      4720      4730      4740      4750
 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCG
4650      4660      4670      4680      4690      4700      4710      4720

     4760      4770      4780      4790      4800      4810      4820
 TTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAG
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAG
       4730      4740      4750      4760      4770      4780      4790

   4830      4840      4850      4860      4870      4880      4890
 CATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCT
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCT
     4800      4810      4820      4830      4840      4850      4860

 4900      4910      4920      4930      4940      4950      4960
 TAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAA
   4870      4880      4890      4900      4910      4920      4930

4970      4980      4990      5000      5010      5020      5030      5040
 CCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 CCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA
 4940      4950      4960      4970      4980      4990      5000

       5050      5060      5070      5080      5090      5100      5110
 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCAT
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCAT
5010      5020      5030      5040      5050      5060      5070      5080

     5120      5130      5140      5150      5160      5170      5180
 TGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAA
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 TGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAA
       5090      5100      5110      5120      5130      5140      5150

   5190      5200      5210      5220      5230 X    5240      5250
 GATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAAT
 ||||||||||||||||||||||||||||||||||||||||||||||||
 GATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTC
     5160      5170      5180      5190      5200

 5260      5270      5280
 TCTGAGGCTTCATCGGGAGCTAGAGC
```

2. US-09-902-432-3 (1-6160)
   US-08-978-277A-1 Sequence 1, Application US/08978277A

Sequence 1, Application US/08978277A
GENERAL INFORMATION:
APPLICANT: Gelman, Irwin H.
TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
STREET: 30 Rockefeller Plaza

```
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  1:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  5074 base pairs
      TYPE:  nucleic acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  cDNA
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:
    ORIGINAL SOURCE:

Initial Score    =   2451  Optimized Score  =    4579  Significance = -0.71
Residue Identity =    90%  Matches          =    4692  Mismatches   =   313
Gaps             =    153  Conservative Substitutions             =     0

      970       980       990      1000      1010      1020      1030
  GCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCG
                                                    ||||||||||||||||||||||
                                                    GGAAAAGACAGAGCCAGCCTCG
                                                    X        10        20

   1040      1050      1060      1070      1080      1090      1100
  GAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCT
        30        40        50        60        70        80        90

 1110      1120      1130      1140      1150      1160      1170      1180
  TTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  TTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT
      100       110       120       130       140       150       160

       1190      1200      1210      1220      1230      1240      1250
  GATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAG
   170       180       190       200       210       220       230

      1260      1270      1280      1290      1300      1310      1320
  GAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAG
 240       250       260       270       280       290       300       310

     1330      1340      1350      1360      1370      1380      1390
  AAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATG
        320       330       340       350       360       370       380

 1400      1410      1420      1430      1440      1450      1460
  TGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  TGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCA
      390       400       410       420       430       440       450

1470      1480      1490      1500      1510      1520      1530      1540
  GAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAG
    460       470       480       490       500       510       520

        1550      1560      1570      1580      1590      1600      1610
  AAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGA
   530       540       550       560       570       580       590

      1620      1630      1640      1650      1660      1670      1680
  GACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAG
 600       610       620       630       640       650       660       670

    1690      1700      1710      1720      1730      1740      1750
  AGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAG
  |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
  AGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCA-CCCAG
        680       690       700       710       720       730       740

 1760      1770      1780      1790      1800      1810      1820
  GATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCC
  |||||||||||||||||||||||||||||   ||||||||||||||| ||| ||||||||||||||||||||
  GATGGGGAAGCTGAGGAAGGAACTACTTC--GTGGAGAGAAGAAGAG-GAA-GGATCACTCCCTGGGCATCC
      750       760       770       780       790       800

1830      1840      1850      1860      1870      1880      1890      1900
  TTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAG
  |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
  TTCAAAAAGATGGTGACACCCAAGAAAC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAG
810       820       830       840       850       860       870       880

       1910      1920      1930      1940      1950      1960      1970
  AAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTT
      890       900       910       920       930       940       950

     1980      1990      2000      2010      2020      2030      2040
  GGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGATTTGT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGATTTGT
     960       970       980       990      1000      1010      1020

   2050      2060      2070      2080      2090      2100      2110
  GTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGA
  ||||||||||||||||||||||||||||||||||||||||||||| || ||| |||||||||||||||||||
  GTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGAT-AT-AAG-AGGGCCAAGGACACTGGGA
   1030      1040      1050      1060      1070      1080      1090

 2120      2130      2140      2150      2160      2170      2180
  -GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCAC
   |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
  GGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAGCC-GAACAGACGCTGTTCCTGCCAGCAC
    1100      1110      1120      1130      1140      1150      1160

2190      2200      2210      2220      2230      2240      2250      2260
  CCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGT
```

```
    1170      1180      1190      1200      1210      1220      1230

       2270      2280      2290      2300      2310      2320
  CTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGA---A
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||   |
  CTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA
  1240      1250      1260      1270      1280      1290      1300

2330        2340      2350      2360      2370      2380      2390
  AGCC-GAAG-ACTCTAG-TGTA-GAGCA-GTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTT
  |||| |||| ||||||| |||| ||||| |||||||||||||||||||||| ||||||||||||||||||||
  AGCCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGAATCTTGGGTT
 1310      1320      1330      1340      1350      1360      1370      1380

  2400      2410      2420      2430      2440      2450      2460
  TCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA-GCAAGAACAAGCCACTGTGGA
  |||||||||||||||||||||||||||||||||||||||||||| ||||| |||||||||||||||||||||
  TCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAAGGCAAGAACAAGCCACTGTGGA
      1390      1400      1410      1420      1430      1440      1450

 2470      2480      2490      2500      2510      2520      2530
  AGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAA
  ||||||||||||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||||| |
  AGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTGATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGA
     1460      1470      1480      1490      1500      1510      1520

2540      2550      2560      2570      2580      2590      2600      2610
  TGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGT
  ||||||||||||||||||||||||||||||||||||| |||||||| |||||||||| ||||||||||| ||
  TGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATGCGGAGCTG-CCCAGCTGCT-GGGGCTGTGTA-GT
    1530      1540      1550      1560      1570      1580      1590

     2620      2630      2640      2650      2660      2670      2680
  GTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCAC
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCAC
    1600      1610      1620      1630      1640      1650      1660

     2690      2700      2710      2720      2730      2740      2750
  CAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGA
    1670      1680      1690      1700      1710      1720      1730

   2760      2770      2780      2790      2800      2810      2820
  AGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCACCCAGAC
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCACCCAGAC
 1740      1750      1760      1770      1780      1790      1800

 2830      2840      2850      2860      2870      2880      2890
  GTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGAC
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGAC
1810      1820      1830      1840      1850      1860      1870      1880

2900      2910      2920      2930      2940      2950      2960      2970
  AGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGA
      1890      1900      1910      1920      1930      1940      1950

     2980      2990      3000      3010      3020      3030      3040
  CATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGT
    1960      1970      1980      1990      2000      2010      2020

    3050      3060      3070      3080      3090      3100      3110
  AGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAA
  |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AGAGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAA
    2030      2040      2050      2060      2070      2080      2090

   3120      3130      3140      3150      3160      3170      3180
  GGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGT
 2100      2110      2120      2130      2140      2150      2160

 3190      3200      3210      3220      3230      3240      3250
  TGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  TGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGT
2170      2180      2190      2200      2210      2220      2230      2240

3260      3270      3280      3290      3300      3310      3320      3330
  GCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGT
      2250      2260      2270      2280      2290      2300      2310

      3340      3350      3360      3370      3380      3390      3400
  TCCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACA
  |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  TCCTGAAGTCACAGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACA
    2320      2330      2340      2350      2360      2370      2380

    3410      3420      3430      3440      3450      3460      3470
  GGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTC
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTC
   2390      2400      2410      2420      2430      2440      2450

   3480      3490      3500      3510      3520      3530      3540
  AGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAG
2460      2470      2480      2490      2500      2510      2520

3550      3560      3570      3580      3590      3600      3610
  GCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGT
2530      2540      2550      2560      2570      2580      2590      2600

3620      3630      3640      3650      3660      3670      3680      3690
  TCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGC
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
  TCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGC
      2610      2620      2630      2640      2650      2660      2670

      3700      3710      3720      3730      3740      3750      3760
  TGCAGCCGTGCCCGTTCTGGC-AAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAA
  |||||||||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
  TGCAGCCGTGCCCG-TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAA
    2680      2690      2700      2710      2720      2730      2740

     3770      3780      3790      3800      3810      3820      3830
  AAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATG
    2750      2760      2770      2780      2790      2800      2810

   3840      3850      3860      3870      3880      3890      3900
  ACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  ACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGG
  2820      2830      2840      2850      2860      2870      2880

  3910      3920      3930      3940      3950      3960      3970
  AGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTG
2890      2900      2910      2920      2930      2940      2950      2960
```

```
3980      3990      4000      4010      4020      4030      4040      4050
  AGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCT
  |||||||||||||||||||||||||||||| | | |||||||||||||||||||||||||||||||||||||
  AGCAGGAAACAGCCGCTCCTGAGCATGAAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCT
      2970      2980      2990      3000      3010      3020      3030

      4060      4070      4080      4090      4100      4110      4120
  CAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCA
      3040      3050      3060      3070      3080      3090      3100

      4130      4140      4150      4160      4170      4180      4190
  TAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  TAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAA
    3110      3120      3130      3140      3150      3160      3170

    4200      4210      4220      4230      4240      4250      4260
  CGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGAAGTCCT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGAAGTCCT
  3180      3190      3200      3210      3220      3230      3240

  4270      4280      4290      4300      4310      4320      4330
  CTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAG
3250      3260      3270      3280      3290      3300      3310      3320

4340      4350      4360      4370      4380      4390      4400      4410
  AATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCG
      3330      3340      3350      3360      3370      3380      3390

      4420      4430      4440      4450      4460      4470      4480
  CATCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CATCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTA
    3400      3410      3420      3430      3440      3450      3460

    4490      4500      4510      4520      4530      4540      4550
  CAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGC
  ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CAGCAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGC
    3470      3480      3490      3500      3510      3520      3530

    4560      4570      4580      4590      4600      4610      4620
  TGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTTATGATT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  TGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTTATGATT
  3540      3550      3560      3570      3580      3590      3600

  4630      4640      4650      4660      4670      4680      4690
  CACAGACCCAGGTTCCTGC-ATGCAGGCTT-GACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGA
  |||||||||||||||||||  ||||  |||| |  ||||||||||||||||||||||||||||||||||||| 
  CACAGACCCAGGTTCCTGCAATGC--GCTTGGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGT
3610      3620      3630      3640      3650      3660      3670

  4700      4710      4720      4730      4740      4750      4760
  TGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGC
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
  TGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCAT-GGC
3680      3690      3700      3710      3720      3730      3740

4770      4780      4790      4800      4810      4820      4830      4840
  TCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCT
  |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  TCAG-CTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCT
3750      3760      3770      3780      3790      3800      3810      3820

      4850      4860      4870      4880      4890      4900      4910

  CCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGC
  ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
  CCTCAACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGG-
    3830      3840      3850      3860      3870      3880      3890

    4920      4930      4940      4950      4960      4970      4980
  CAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCGAT-C
  |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||| |
  CAGTG--TCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGA--GCGATGC
      3900      3910      3920      3930      3940      3950

    4990      5000      5010      5020      5030      5040      5050
  CCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  CCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAG
3960      3970      3980      3990      4000      4010      4020      4030

  5060      5070      5080      5090      5100      5110      5120
  AGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGA
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  AGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGA
      4040      4050      4060      4070      4080      4090      4100

5130      5140      5150      5160      5170      5180      5190
  CCAGAATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCAT
  ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
  CCAGAATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCAT
    4110      4120      4130      4140      4150      4160      4170

5200      5210      5220      5230      5240      5250      5260      5270
  GAGATCCAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCG
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GAGATCCAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCG
    4180      4190      4200      4210      4220      4230      4240

      5280      5290      5300      5310      5320      5330      5340
  GGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTAT
  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  GGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTAT
    4250      4260      4270      4280      4290      4300      4310

        5350      5360      5370      5380      5390      5400
  TTC---TAACATTTCCT-CGTT-TCAAGACTGCCTTTGATT-TGCCCCTTGATGCC--GTCCGTGTATTTCG
  |||     |     |||| ||||  |||          ||| ||||        || ||  |  |   | ||
  TTCGGATTTAAGGTCCTGCGTTCTCAA-CCTGGAACCAATTCTGCC-----ATACCTAGTTCCACT-TCTCA
4320      4330      4340      4350      4360      4370      4380

5410      5420      5430      5440      5450      5460      5470
  GATTTAAG-GTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCCACTTCTCAAACTGGAGCA
   | |  ||  ||||  | ||  |   |    |     |  ||   || ||||  |  || ||  | ||   | |
  AACTGGAGCATCCTCCTTTATGTATTT----ATATGTATGTTTATGTAGTCCTCCTCCTGTACCTATTGTA
    4390      4400      4410      4420      4430      4440      4450

5480      5490      5500      5510      5520      5530      5540
  TCCTCCTTTATGTA---TTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCT
  |  |  ||| | ||     ||||      ||| ||| |||| |    |     | || |   |||  |   | |
  TATT--TTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCG
      4460      4470      4480      4490      4500      4510      4520

5550      5560      5570      5580      5590      5600
  AACGTTT---AAGC-ACATGCTT--TTTGTATTATGC----AATATATAACGGGTGT-GCAGCCATAG-CGA
  |   |||    ||||  || || |    ||||   |||     || | |||||     | |||   | | | |
  ACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAA
      4530      4540      4550      4560      4570      4580      4590

5610      5620      5630      5640      5650      5660
  CGCT-TTGAAAAGCT--CCAAGCCT--CAACTGTAACCTGC-AG------CAAACAGATAACATTCCTGGCA
   || ||  ||||  |   |   |   | ||   | ||||   || | ||       |||  |  |   ||||  |
  GTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTCCT---A
    4600      4610      4620      4630      4640      4650      4660

5670      5680      5690      5700      5710      5720      5730
  AGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCC---TCTGAAAGTGG
```

```
  | | || |  ||     ||||    |  || | | |  |   || |     ||||   | | |   ||
TGCACAGCGAGCTCAGAAATAAA--AACCCCATTTTGAAACATCCAGGAT----GTCCCAATATTACCATGA
     4670      4680      4690      4700      4710      4720

5740      5750      5760      5770      5780      5790      5800
TTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACA--TCCAGGATGTCCCAATATT
|| ||||||    | |     |||    |||    |  ||   || |||| |  ||      ||| || |||
TT-TTTTCC---CCCCTTTTTGCT----AAT-CCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCA-GATT
4730      4740      4750      4760      4770      4780

  5810      5820      5830      5840      5850      5860      5870
A--CCATGATTTTTCCCCCCTTTTGCTAATCCAGT---CC-AGGTTGGAAAGAAGTCTCCTCTGTGTCAG
|  || ||   | | |      | |   | |||  |||    || | |      ||| || ||||    || |||
AAGCCCTG--TCTCTTAATGATATGGACAAAT-GAGTGTGCCTAAGGCCATGAGATGTTTCCT-AATG-CAG
4790      4800      4810      4820      4830      4840      4850

   5880      5890      5900      5910      5920      5930      5940
ATTAAGCCCTGTCTCTTAATGATATGGACAAATG-AGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGA
|   |   |||| | |   |  | | |   | || | | |  ||| |   |  ||     ||| | ||||||
AAGGA-ATCTGT-TGTACGTTTTTTTG---ATTGTACTCT-TCTATG--CTGGACCGAATTCAT-ATGCAGA
4860      4870      4880      4890      4900      4910

  5950      5960      5970      5980      5990      6000      6010
AGGAATCTGTTGTAC-GTTTTTT--TGATTGTACTCT--TCTATGCTGGACCGAATTCATATGCAGATCGAA
 |||  ||  || | ||| |||   ||| ||| | |   |  || |||     || ||  | ||  | |  |
TCGAA-GTG-AGTCCTGTTCTTTACAGATGGTATTTGATAGATACTG----GAGTTTGTCTG-TGTTATAT
4920      4930      4940      4950      4960      4970      4980

   6020      6030      6040      6050      6060      6070      6080
GTGAGTCCTGTTCTTTACAGATGGTATTTGATAGATACTGGAGTTTGTCTGTGTTAT-ATCTGTGCCCCTT
 || | | ||  ||||||| |||   || |||    ||  |    ||||  |   || | || ||  |  ||
CTGTG-CCCCTTCTTTA-AGA-ACAATGTTGCATTATGTT--CCTTTGGATAAATTGTGATTTG-ACAACTG
   4990      5000      5010      5020      5030      5040      5050

   6090      6100    X 6110      6120      6130      6140      6150
CTTT-AAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATAT
 ||| || ||   || ||  | ||
ATTTAAATAAACATATTTGACTAC
   5060      5070    X
```

TT

```
 > O <
O| |O  IntelliGenetics
 > O <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file us-09-902-432-4.res made by bshears on Fri 13 Dec 102 15:27:21-PST.


Query sequence being compared:US-09-902-432-4 (1-1596)
Number of sequences searched:                 18
Number of scores above cutoff:                18

      Results of the initial comparison of US-09-902-432-4 (1-1596) with:
        File : /home/bshears/*9782*pep

   100-
      -
N     -
U   50-
M     -
B     -
E     -
R     *
      -
O     -
F   10-
      -
S     -
E    5-
Q     -
U     -
E     -
N     -
C     -
E     -                               *                                  *
S    0--------------------------------------------------------------------------
      |      |      |      |      |  |    |      |    |  |      |      ||
SCORE 0     177    355    532    709 |   887   1064   1241   1419   1596
STDEV                1                  2                 3               4


                                PARAMETERS

Similarity matrix       Unitary    K-tuple                      2
Mismatch penalty              1    Joining penalty             20
Gap penalty                1.00    Window size                 32
Gap size penalty           0.05
Cutoff score                  0
Randomization group           0

                                SEARCH STATISTICS

Scores:                  Mean        Median      Standard Deviation
                          136           8         399.84

Times:                   CPU                     Total Elapsed
                     00:00:00.00                 00:00:00.00

Number of residues:                 3203
Number of sequences searched:         18
Number of scores above cutoff:        18


The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was found:

                                                     Init. Opt.
Sequence Name     Description                 Length Score Score  Sig. Frame
------------------------------------------------------------------------------
   1. US-08-978-277A-4 Sequence 4, Application U   1596  1596  1596   3.65   0

The list of other best scores is:

                                                     Init. Opt.
Sequence Name     Description                 Length Score Score  Sig. Frame
------------------------------------------------------------------------------
                  ****  1 standard deviation above mean ****
   2. US-08-978-277A-2 Sequence 2, Application U   1346   710  1094   1.44   0
                  ****  0 standard deviation from mean ****
   3. US-08-978-277A-1 Sequence 18, Application      23    23    23  -0.28   0
   4. US-08-978-277A-1 Sequence 17, Application      23    23    23  -0.28   0
   5. US-08-978-277A-1 Sequence 16, Application      23    23    23  -0.28   0
   6. US-08-978-277A-1 Sequence 15, Application      23    13    22  -0.31   0
   7. US-08-978-277A-7 Sequence 7, Application U     10    10    10  -0.32   0
   8. US-08-978-277A-2 Sequence 20, Application      19     9    11  -0.32   0
   9. US-08-978-277A-1 Sequence 19, Application       7     7     7  -0.32   0
  10. US-08-978-277A-1 Sequence 13, Application      17     7     8  -0.32   0
  11. US-08-978-277A-1 Sequence 12, Application      23     7    10  -0.32   0
  12. US-08-978-277A-1 Sequence 11, Application      23     7    10  -0.32   0
  13. US-08-978-277A-8 Sequence 8, Application U     10     5     5  -0.33   0
  14. US-08-978-277A-6 Sequence 6, Application U     10     5     5  -0.33   0
  15. US-08-978-277A-5 Sequence 5, Application U     10     4     4  -0.33   0
  16. US-08-978-277A-1 Sequence 14, Application      20     4     4  -0.33   0
  17. US-08-978-277A-9 Sequence 9, Application U     10     3     4  -0.33   0
  18. US-08-978-277A-1 Sequence 10, Application      10     2     2  -0.34   0


1. US-09-902-432-4 (1-1596)
   US-08-978-277A-4 Sequence 4, Application US/08978277A

 Sequence 4, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  4:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  1596 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
```

ORIGINAL SOURCE:

```
Initial Score     =   1596  Optimized Score  =   1596  Significance =  3.65
Residue Identity  =   100%  Matches          =   1596  Mismatches   =     0
Gaps              =      0  Conservative Substitutions              =     0

       X        10        20        30        40        50        60        70
       MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSSVNGVAEQGDVHV
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSSVNGVAEQGDVHV
       X        10        20        30        40        50        60        70

             80        90       100       110       120       130       140
       QEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQP
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       QEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQP
             80        90       100       110       120       130       140

         150       160       170       180       190       200       210
       AESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKE
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       AESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKE
         150       160       170       180       190       200       210

        220       230       240       250       260       270       280
       SELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFF
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       SELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFF
        220       230       240       250       260       270       280

      290       300       310       320       330       340       350       360
       THGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       THGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL
      290       300       310       320       330       340       350       360

             370       380       390       400       410       420       430
       EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEK
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEK
             370       380       390       400       410       420       430

          440       450       460       470       480       490       500
       LAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKK
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       LAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKK
          440       450       460       470       480       490       500

         510       520       530       540       550       560       570
       LFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQD
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       LFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQD
         510       520       530       540       550       560       570

        580       590       600       610       620       630       640
       GEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       GEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVG
        580       590       600       610       620       630       640

      650       660       670       680       690       700       710       720
       EEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       EEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ
      650       660       670       680       690       700       710       720

             730       740       750       760       770       780       790
       EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKK
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKK
             730       740       750       760       770       780       790

          800       810       820       830       840       850       860
       FIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEEL
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       FIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEEL
          800       810       820       830       840       850       860

        870       880       890       900       910       920       930
       SKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       SKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEG
        870       880       890       900       910       920       930

       940       950       960       970       980       990      1000
       KDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGV
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       KDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGV
       940       950       960       970       980       990      1000

     1010      1020      1030      1040      1050      1060      1070      1080
       LDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       LDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG
     1010      1020      1030      1040      1050      1060      1070      1080

            1090      1100      1110      1120      1130      1140      1150
       AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTAD
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTAD
            1090      1100      1110      1120      1130      1140      1150

         1160      1170      1180      1190      1200      1210      1220
       GTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVP
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       GTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVP
         1160      1170      1180      1190      1200      1210      1220

        1230      1240      1250      1260      1270      1280      1290
       VLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEME
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       VLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEME
        1230      1240      1250      1260      1270      1280      1290

       1300      1310      1320      1330      1340      1350      1360
       TDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQV
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       TDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQV
       1300      1310      1320      1330      1340      1350      1360

     1370      1380      1390      1400      1410      1420      1430      1440
       QSLDTTVTQTAEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       QSLDTTVTQTAEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI
     1370      1380      1390      1400      1410      1420      1430      1440

            1450      1460      1470      1480      1490      1500      1510
       IVTPAPESTLHPDLQGEISASQRERSEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQ
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       IVTPAPESTLHPDLQGEISASQRERSEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQ
            1450      1460      1470      1480      1490      1500      1510

          1520      1530      1540      1550      1560      1570      1580
       TAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAWAQPRKC
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
       TAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAWAQPRKC
          1520      1530      1540      1550      1560      1570      1580

        1590      X
       LPRLQLKAPVSK
       ||||||||||||
       LPRLQLKAPVSK
        1590      X
```

2. US-09-902-432-4 (1-1596)

```
US-08-978-277A-2 Sequence 2, Application US/08978277A

Sequence 2, Application US/08978277A
 GENERAL INFORMATION:
   APPLICANT:  Gelman, Irwin H.
   TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
   NUMBER OF SEQUENCES:  20
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
     STREET:  30 Rockefeller Plaza
     CITY:  New York
     STATE:  NY
     COUNTRY:  USA
     ZIP:  10112-0228
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Diskette
     COMPUTER:  IBM Compatible
     OPERATING SYSTEM:  DOS
     SOFTWARE:  FastSEQ Version 1.5
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/978,277A
     FILING DATE:
     CLASSIFICATION:  514
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  08/665,401
     FILING DATE:  18-JUN-1996
   ATTORNEY/AGENT INFORMATION:
     NAME:  Clark, Richard S
     REGISTRATION NUMBER:  26,154
     REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  212-408-2558
     TELEFAX:  212-765-2519
     TELEX:
 INFORMATION FOR SEQ ID NO:  2:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  1346 amino acids
     TYPE:  amino acid
     STRANDEDNESS:  single
     TOPOLOGY:  linear
   MOLECULE TYPE:  peptide
   HYPOTHETICAL:  NO
   ANTI-SENSE:  NO
   FRAGMENT TYPE:  internal
   ORIGINAL SOURCE:

Initial Score    =    710  Optimized Score  =   1094  Significance =  1.44
Residue Identity =    89%  Matches          =   1106  Mismatches   =    86
Gaps             =     37  Conservative Substitutions              =     0

     340       350       360       370       380       390       400
EQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEE
                                                  ||||||||||||||||||||||
                                                  MEAHQEVVAEVHVSTVEKTEEE
                                                  X        10        20

   410       420       430       440       450       460       470       480
QGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
QGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE
      30        40        50        60        70        80        90

        490       500       510       520       530       540       550
GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGES
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGES
    100       110       120       130       140       150       160

      560       570       580       590       600       610       620
SASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEK
||||||||||||||||||||||  |          ||||        |   |    ||||||||||||||||
SASSPEEPEETTCLEKGPLEAPRMG-KLRKELLRGEKKRKDHSLGILQKDGDT-QETVRRPSESDKEEELEK
  170       180       190       200       210       220       230

     630       640       650       660       670       680       690
VKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGG
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |  |||||||
VKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSS-DIRGPRTLGG
  240       250       260       270       280       290       300

  700       710       720       730       740       750       760
DSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEK-A
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
GQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKEA
310       320       330       340       350       360       370

  770          780             790       800       810       820
EDSSV-----EQLSTEIE-----PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAV
    |        |   |     | |  |         ||  |   | ||||||||||||||||| | ||||
GRTLVVGAGCPLRSNRVEKNLGFPLRNSS------PDGGR--KGQMGRQEQATVEDSGPVEINEDEPDVPAV
380       390       400                 410       420       430       440

830       840       850       860       870       880       890       900
VPLSEYNAVEREKMEAQGNTELPQLLGAVYYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEP
|||||| |||||||||||| |||    | |  |||||||||||||||||||||||||||||||||||||||
VPLSEYDAVEREKMEAQGNAELPSCWGCV-VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEP
      450       460       470       480       490       500       510

      910       920       930       940       950       960       970
LEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEAS
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEAS
    520       530       540       550       560       570       580

     980       990      1000      1010      1020      1030      1040
GAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTG
||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
GAEETTDMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTG
  590       600       610       620       630       640       650

  1050      1060      1070      1080      1090      1100      1110
SKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIK
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
SKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIK
660       670       680       690       700       710       720       730

1120      1130      1140      1150      1160      1170      1180
LQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEP
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEP
       740       750       760       770       780       790       800

1190      1200      1210      1220      1230      1240      1250      1260
STLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQK
|||||||||||||||||||||||||||| ||||||| | |||||||||||||||||||||||||||||||||
STLPNNVPAQEEHGEEPGRDVLEPTQQELAAAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQK
    810       820       830       840       850       860       870

     1270      1280      1290      1300      1310      1320      1330
AADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVL
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||||||
AADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVL
    880       890       900       910       920       930       940

    1340      1350      1360      1370      1380      1390      1400
TLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAHL
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
TLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAHL
  950       960       970       980       990      1000      1010

  1410      1420      1430      1440      1450      1460      1470
LPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPD
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
LPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPD
1020      1030      1040      1050      1060      1070      1080      1090
```

```
 1480      1490      1500      1510      1520      1530      1540
  ADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRC
  |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||
  ADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPAMRLDSREPNRC
        1100      1110      1120      1130      1140      1150      1160

1550      1560      1570      1580      1590          X
  WTKMKDAKMKHPVPQPREDLQVLTVLEAWAQPRKCLPRLQLKA--------PVSK
  ||||| |||||||||||||||||||||        |    | |         ||
  WTKMKVAKMKHPVPQPREDLQVLTVLEAW------LSSEMLAALAVESAGVKVSIEKLPPQPKDQKEHAADG
      1170      1180      1190              1200      1210      1220

  PQLQSLAQAEAVSGNLTKESPDTNGPKLTEERC
1230      1240      1250      1260
```

3. US-09-902-432-4 (1-1596)
US-08-978-277A-1 Sequence 18, Application US/08978277A

```
 Sequence 18, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  18:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  23 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score     =     23  Optimized Score  =      23  Significance = -0.28
Residue Identity =    100%  Matches          =      23  Mismatches    =     0
Gaps              =      0  Conservative Substitutions                =     0

        700       710       720       730       740       750       760
  PRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSK
                                                      |||||||||||||||||||||||
                                                      EGVSTWESFKRLVTPRKKSKSK
                                                      X        10        20

  X    770       780       790       800       810
  LEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVED
  |
  L
  X
```

4. US-09-902-432-4 (1-1596)
US-08-978-277A-1 Sequence 17, Application US/08978277A

```
 Sequence 17, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  17:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  23 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score     =     23  Optimized Score  =      23  Significance = -0.28
Residue Identity =    100%  Matches          =      23  Mismatches    =     0
Gaps              =      0  Conservative Substitutions                =     0

        550       560       570       580       590 X     600       610
  PESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRP
                                                    ||||||||||||||||||||||
                                                    EGITPWASFKKMVTPKKRVRRP
                                                    X        10        20

  X   620       630       640       650       660
  SESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSV
  |
  S
  X
```

5. US-09-902-432-4 (1-1596)
US-08-978-277A-1 Sequence 16, Application US/08978277A

```
Sequence 16, Application US/08978277A
 GENERAL INFORMATION:
   APPLICANT:  Gelman, Irwin H.
   TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
   NUMBER OF SEQUENCES:  20
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
     STREET:  30 Rockefeller Plaza
     CITY:  New York
     STATE:  NY
     COUNTRY:  USA
     ZIP:  10112-0228
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Diskette
     COMPUTER:  IBM Compatible
     OPERATING SYSTEM:  DOS
     SOFTWARE:  FastSEQ Version 1.5
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/978,277A
     FILING DATE:
     CLASSIFICATION:  514
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  08/665,401
     FILING DATE:  18-JUN-1996
   ATTORNEY/AGENT INFORMATION:
     NAME:  Clark, Richard S
     REGISTRATION NUMBER:  26,154
     REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  212-408-2558
     TELEFAX:  212-765-2519
     TELEX:
 INFORMATION FOR SEQ ID NO:  16:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  23 amino acids
     TYPE:  amino acid
     STRANDEDNESS:  single
     TOPOLOGY:  linear
   MOLECULE TYPE:  peptide
   HYPOTHETICAL:  NO
   ANTI-SENSE:  NO
   FRAGMENT TYPE:  internal
   ORIGINAL SOURCE:

Initial Score    =     23  Optimized Score  =     23  Significance = -0.28
Residue Identity =   100%  Matches          =     23  Mismatches   =     0
Gaps             =      0  Conservative Substitutions              =     0

       460       470       480       490       500   X   510       520
    EMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGG
                                                      |||||||||||||||||||||
                                                      KLFSSSGLKKLSGKKQKGKRGG
                                                      X        10        20

    X 530       540       550       560       570
    GGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQD
    |
    G
    X


6. US-09-902-432-4 (1-1596)
   US-08-978-277A-1 Sequence 15, Application US/08978277A

Sequence 15, Application US/08978277A
 GENERAL INFORMATION:
   APPLICANT:  Gelman, Irwin H.
   TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
   NUMBER OF SEQUENCES:  20
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
     STREET:  30 Rockefeller Plaza
     CITY:  New York
     STATE:  NY
     COUNTRY:  USA
     ZIP:  10112-0228
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Diskette
     COMPUTER:  IBM Compatible
     OPERATING SYSTEM:  DOS
     SOFTWARE:  FastSEQ Version 1.5
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/978,277A
     FILING DATE:
     CLASSIFICATION:  514
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  08/665,401
     FILING DATE:  18-JUN-1996
   ATTORNEY/AGENT INFORMATION:
     NAME:  Clark, Richard S
     REGISTRATION NUMBER:  26,154
     REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  212-408-2558
     TELEFAX:  212-765-2519
     TELEX:
 INFORMATION FOR SEQ ID NO:  15:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  23 amino acids
     TYPE:  amino acid
     STRANDEDNESS:  single
     TOPOLOGY:  linear
   MOLECULE TYPE:  peptide
   HYPOTHETICAL:  NO
   ANTI-SENSE:  NO
   FRAGMENT TYPE:  internal
   ORIGINAL SOURCE:

Initial Score    =     13  Optimized Score  =     22  Significance = -0.31
Residue Identity =    76%  Matches          =     23  Mismatches   =     0
Gaps             =      7  Conservative Substitutions              =     0

   230       240       250       260       270       280       290       300
    TLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS
                                                      ||||||||||||       ||
                                                      ETTSSFKKFFTHG-------TS
                                                      X        10

          310       320       330       340       350
    FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVEL
    ||||||||
    FKKSKEDD
      20  X


7. US-09-902-432-4 (1-1596)
   US-08-978-277A-7 Sequence 7, Application US/08978277A

Sequence 7, Application US/08978277A
 GENERAL INFORMATION:
   APPLICANT:  Gelman, Irwin H.
   TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
   NUMBER OF SEQUENCES:  20
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
     STREET:  30 Rockefeller Plaza
     CITY:  New York
     STATE:  NY
     COUNTRY:  USA
     ZIP:  10112-0228
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Diskette
     COMPUTER:  IBM Compatible
     OPERATING SYSTEM:  DOS
     SOFTWARE:  FastSEQ Version 1.5
```

```
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER:  US/08/978,277A
    FILING DATE:
    CLASSIFICATION:  514
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER:  08/665,401
    FILING DATE:  18-JUN-1996
  ATTORNEY/AGENT INFORMATION:
    NAME:  Clark, Richard S
    REGISTRATION NUMBER:  26,154
    REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
  TELECOMMUNICATION INFORMATION:
    TELEPHONE:  212-408-2558
    TELEFAX:  212-765-2519
    TELEX:
INFORMATION FOR SEQ ID NO:  7:
  SEQUENCE CHARACTERISTICS:
    LENGTH:  10 amino acids
    TYPE:  amino acid
    STRANDEDNESS:  single
    TOPOLOGY:  linear
  MOLECULE TYPE:  peptide
  HYPOTHETICAL:  NO
  ANTI-SENSE:  NO
  FRAGMENT TYPE:  internal
  ORIGINAL SOURCE:

Initial Score    =     10  Optimized Score  =     10  Significance = -0.32
Residue Identity =   100%  Matches          =     10  Mismatches   =     0
Gaps             =      0  Conservative Substitutions              =     0

    X        X        20        30        40        50        60
    MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS
    ||||||||||
    MGAGSSTEQR
    X        10

8. US-09-902-432-4 (1-1596)
   US-08-978-277A-2 Sequence 20, Application US/08978277A

 Sequence 20, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
```

```
  INFORMATION FOR SEQ ID NO:  20:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  19 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score    =      9  Optimized Score  =     11  Significance = -0.32
Residue Identity =    60%  Matches          =     12  Mismatches   =     7
Gaps             =      1  Conservative Substitutions              =     0

          550       560       570       580       590 X     600       610
    PESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRP
                                                      ||   | |||| |||||
                                                      EG-XXWXSFKKXVTPKKKXK
                                                      X        10        X

        620       630       640       650       660
    SESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVD

9. US-09-902-432-4 (1-1596)
   US-08-978-277A-1 Sequence 19, Application US/08978277A

 Sequence 19, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  19:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  7 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:
```

```
Initial Score     =      7 Optimized Score  =       7 Significance = -0.32
Residue Identity =   100% Matches          =       7 Mismatches    =     0
Gaps              =      0 Conservative Substitutions            =     0

       250       260       270       280       290 X     300       310
 QAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAE
                                                  |||||||
                                                  WAGWRKK
                                                  X     X

       320       330       340
 KRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQAR
```

10. US-09-902-432-4 (1-1596)
    US-08-978-277A-1 Sequence 13, Application US/08978277A

```
 Sequence 13, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  13:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  17 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score     =      7 Optimized Score  =       8 Significance = -0.32
Residue Identity =    47% Matches          =       8 Mismatches    =     9
Gaps              =      0 Conservative Substitutions            =     0

       460       470       480       490       500  X  510       520
 EMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGG
                                                  | |||    ||||
                                                  KKFSSKKPFKLSGFSFR
                                                  X        10       X

     530       540       550       560       570
 GGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGP
```

11. US-09-902-432-4 (1-1596)
    US-08-978-277A-1 Sequence 12, Application US/08978277A

```
 Sequence 12, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  12:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  23 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score     =      7 Optimized Score  =      10 Significance = -0.32
Residue Identity =    47% Matches          =      11 Mismatches    =     9
Gaps              =      3 Conservative Substitutions            =     0

       460       470       480       490       500  X  510       520
 EMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSG---KKQKGK
                                                  | |||    ||||   || |
                                                  KRFSSKKSFKLSGFSFKKSKKE
                                                  X        10        20

 X   530       540       550       560       570
 RGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEA

 A
 X
```

12. US-09-902-432-4 (1-1596)
    US-08-978-277A-1 Sequence 11, Application US/08978277A

```
 Sequence 11, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
```

```
  NUMBER OF SEQUENCES:  20
  CORRESPONDENCE ADDRESS:
    ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
    STREET:  30 Rockefeller Plaza
    CITY:  New York
    STATE:  NY
    COUNTRY:  USA
    ZIP:  10112-0228
  COMPUTER READABLE FORM:
    MEDIUM TYPE:  Diskette
    COMPUTER:  IBM Compatible
    OPERATING SYSTEM:  DOS
    SOFTWARE:  FastSEQ Version 1.5
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER:  US/08/978,277A
    FILING DATE:
    CLASSIFICATION:  514
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER:  08/665,401
    FILING DATE:  18-JUN-1996
  ATTORNEY/AGENT INFORMATION:
    NAME:  Clark, Richard S
    REGISTRATION NUMBER:  26,154
    REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
  TELECOMMUNICATION INFORMATION:
    TELEPHONE:  212-408-2558
    TELEFAX:  212-765-2519
    TELEX:
INFORMATION FOR SEQ ID NO:  11:
  SEQUENCE CHARACTERISTICS:
    LENGTH:  23 amino acids
    TYPE:  amino acid
    STRANDEDNESS:  single
    TOPOLOGY:  linear
  MOLECULE TYPE:  peptide
  HYPOTHETICAL:  NO
  ANTI-SENSE:  NO
  FRAGMENT TYPE:  internal
  ORIGINAL SOURCE:

Initial Score     =       7  Optimized Score  =     10  Significance = -0.32
Residue Identity =     47%  Matches          =     11  Mismatches    =     9
Gaps              =       3  Conservative Substitutions           =     0

        460        470        480        490        500    X   510        520
    EMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSG---KKQKGK
                                                      | |||    ||||   || |
                                                      KRFSSKKSFKLSGFSFKKNKKE
                                                      X       10         20

    X    530        540        550        560        570
    RGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEA

    A
    X

13. US-09-902-432-4 (1-1596)
    US-08-978-277A-8 Sequence 8, Application US/08978277A

 Sequence 8, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  8:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  10 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score     =       5  Optimized Score  =      5  Significance = -0.33
Residue Identity =     50%  Matches          =      5  Mismatches    =     5
Gaps              =       0  Conservative Substitutions           =     0

        320        330        340        350        360    X  370    X   380
    RKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEK
                                                              || |||
                                                              MGAGASAEEK
                                                              X        10

      390        400        410        420
    MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGT

14. US-09-902-432-4 (1-1596)
    US-08-978-277A-6 Sequence 6, Application US/08978277A

 Sequence 6, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
```

```
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  6:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  10 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score     =      5  Optimized Score  =      5  Significance = -0.33
Residue Identity =     50%  Matches          =      5  Mismatches   =     5
Gaps             =       0  Conservative Substitutions              =     0

   250       260       270       280       290       300       X        320
  AEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEA
                                                            |||||
                                                        MGCIKSKEDK
                                                        X       10

          330       340       350
  EKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVEL
```

15. US-09-902-432-4 (1-1596)
    US-08-978-277A-5 Sequence 5, Application US/08978277A

```
 Sequence 5, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  5:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  10 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score     =      4  Optimized Score  =      4  Significance = -0.33
Residue Identity =     40%  Matches          =      4  Mismatches   =     6
Gaps             =       0  Conservative Substitutions              =     0

        710       720       730       740       750    X    760   X    770
  ASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQL
                                                          ||| |
                                                      MGSSKSKPKD
                                                      X        10

      780       790       800       810
  STEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVED
```

16. US-09-902-432-4 (1-1596)
    US-08-978-277A-1 Sequence 14, Application US/08978277A

```
 Sequence 14, Application US/08978277A
  GENERAL INFORMATION:
    APPLICANT:  Gelman, Irwin H.
    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
    NUMBER OF SEQUENCES:  20
    CORRESPONDENCE ADDRESS:
      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
      STREET:  30 Rockefeller Plaza
      CITY:  New York
      STATE:  NY
      COUNTRY:  USA
      ZIP:  10112-0228
    COMPUTER READABLE FORM:
      MEDIUM TYPE:  Diskette
      COMPUTER:  IBM Compatible
      OPERATING SYSTEM:  DOS
      SOFTWARE:  FastSEQ Version 1.5
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER:  US/08/978,277A
      FILING DATE:
      CLASSIFICATION:  514
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER:  08/665,401
      FILING DATE:  18-JUN-1996
    ATTORNEY/AGENT INFORMATION:
      NAME:  Clark, Richard S
      REGISTRATION NUMBER:  26,154
      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
    TELECOMMUNICATION INFORMATION:
      TELEPHONE:  212-408-2558
      TELEFAX:  212-765-2519
      TELEX:
  INFORMATION FOR SEQ ID NO:  14:
    SEQUENCE CHARACTERISTICS:
      LENGTH:  20 amino acids
      TYPE:  amino acid
      STRANDEDNESS:  single
      TOPOLOGY:  linear
    MOLECULE TYPE:  peptide
    HYPOTHETICAL:  NO
    ANTI-SENSE:  NO
    FRAGMENT TYPE:  internal
    ORIGINAL SOURCE:

Initial Score     =      4  Optimized Score  =      4  Significance = -0.33
Residue Identity =     20%  Matches          =      4  Mismatches   =    16
Gaps             =       0  Conservative Substitutions              =     0

         690       700       710       720       730  X    740       750 X
```

```
SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVT
                                                                ||    ||
                                                      KRRWKKAFIAVSAAARFKKC
                                                      X        10         20

  760        770        780        790        800
PRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRA
```

17. US-09-902-432-4 (1-1596)
US-08-978-277A-9 Sequence 9, Application US/08978277A

```
Sequence 9, Application US/08978277A
 GENERAL INFORMATION:
   APPLICANT:  Gelman, Irwin H.
   TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
   NUMBER OF SEQUENCES:  20
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
     STREET:  30 Rockefeller Plaza
     CITY:  New York
     STATE:  NY
     COUNTRY:  USA
     ZIP:  10112-0228
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Diskette
     COMPUTER:  IBM Compatible
     OPERATING SYSTEM:  DOS
     SOFTWARE:  FastSEQ Version 1.5
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/978,277A
     FILING DATE:
     CLASSIFICATION:  514
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  08/665,401
     FILING DATE:  18-JUN-1996
   ATTORNEY/AGENT INFORMATION:
     NAME:  Clark, Richard S
     REGISTRATION NUMBER:  26,154
     REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  212-408-2558
     TELEFAX:  212-765-2519
     TELEX:
  INFORMATION FOR SEQ ID NO:  9:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  10 amino acids
     TYPE:  amino acid
     STRANDEDNESS:  single
     TOPOLOGY:  linear
   MOLECULE TYPE:  peptide
   HYPOTHETICAL:  NO
   ANTI-SENSE:  NO
   FRAGMENT TYPE:  internal
   ORIGINAL SOURCE:

Initial Score     =      3  Optimized Score  =      4  Significance = -0.33
Residue Identity =    40%  Matches          =      4  Mismatches   =      6
Gaps              =      0  Conservative Substitutions              =      0

     580        590        600        610        620      X 630     X 640
 DGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTV
                                                         |||  |
                                                      MGCTLSAEDK
                                                      X       10

  650        660        670        680
 GEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSS
```

18. US-09-902-432-4 (1-1596)
US-08-978-277A-1 Sequence 10, Application US/08978277A

```
Sequence 10, Application US/08978277A
 GENERAL INFORMATION:
   APPLICANT:  Gelman, Irwin H.
   TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
   NUMBER OF SEQUENCES:  20
   CORRESPONDENCE ADDRESS:
     ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
     STREET:  30 Rockefeller Plaza
     CITY:  New York
     STATE:  NY
     COUNTRY:  USA
     ZIP:  10112-0228
   COMPUTER READABLE FORM:
     MEDIUM TYPE:  Diskette
     COMPUTER:  IBM Compatible
     OPERATING SYSTEM:  DOS
     SOFTWARE:  FastSEQ Version 1.5
   CURRENT APPLICATION DATA:
     APPLICATION NUMBER:  US/08/978,277A
     FILING DATE:
     CLASSIFICATION:  514
   PRIOR APPLICATION DATA:
     APPLICATION NUMBER:  08/665,401
     FILING DATE:  18-JUN-1996
   ATTORNEY/AGENT INFORMATION:
     NAME:  Clark, Richard S
     REGISTRATION NUMBER:  26,154
     REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
   TELECOMMUNICATION INFORMATION:
     TELEPHONE:  212-408-2558
     TELEFAX:  212-765-2519
     TELEX:
  INFORMATION FOR SEQ ID NO:  10:
   SEQUENCE CHARACTERISTICS:
     LENGTH:  10 amino acids
     TYPE:  amino acid
     STRANDEDNESS:  single
     TOPOLOGY:  linear
   MOLECULE TYPE:  peptide
   HYPOTHETICAL:  NO
   ANTI-SENSE:  NO
   FRAGMENT TYPE:  internal
   ORIGINAL SOURCE:

Initial Score     =      2  Optimized Score  =      2  Significance = -0.34
Residue Identity =    20%  Matches          =      2  Mismatches   =      8
Gaps              =      0  Conservative Substitutions              =      0

     560        570        580        590        600  X    610 X      620
 SASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEK
                                                       |      |
                                                       MLCCMRRTKQ
                                                       X        10

   630        640        650        660
 VKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDT
```

```
OM nucleic - nucleic search, using sw model

Run on:          December 12, 2002, 14:13:30 ; Search time 11503 Seconds
                                          (without alignments)
                                          13464.116 Million cell updates/sec

Title:           US-09-902-432-3
Perfect score:   6160
Sequence:        1 cgggccggcgggagtagaag..........aaataaacatatttgactac 6160

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters:      49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Pending_Patents_NA_Main:*
                 1:  /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
                 2:  /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
                 3:  /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
                 4:  /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
                 5:  /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
                 6:  /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
                 7:  /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
                 8:  /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
                 9:  /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
                 10:  /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
                 11:  /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
                 12:  /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
                 13:  /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
                 14:  /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
                 15:  /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
                 16:  /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
                 17:  /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
                 18:  /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
                 19:  /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
                 20:  /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
                 21:  /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
                 22:  /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
                 23:  /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
                 24:  /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
                 25:  /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
                 26:  /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
                 27:  /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
                 28:  /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
                 29:  /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
                 30:  /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
                 31:  /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
                 32:  /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
                 33:  /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
                 34:  /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
                 35:  /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
                 36:  /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
                 37:  /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
                 38:  /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
                 39:  /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
                 40:  /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
                 41:  /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
                 42:  /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
                 43:  /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
                 44:  /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
                 45:  /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
                 46:  /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
                 47:  /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
                 48:  /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
                 49:  /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
                 50:  /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
                 51:  /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
                 52:  /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
                 53:  /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
                 54:  /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
                 55:  /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
                 56:  /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
                 57:  /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
                 58:  /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
                 59:  /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
                 60:  /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
                 61:  /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
                 62:  /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
                 63:  /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
                 64:  /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
                 65:  /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
                 66:  /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
                 67:  /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
                 68:  /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
                 69:  /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
                 70:  /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
                 71:  /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
                 72:  /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
                 73:  /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
                 74:  /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
                 75:  /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
                 76:  /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
                 77:  /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
                 78:  /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
                 79:  /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
                 80:  /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
                 81:  /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
                 82:  /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
                 83:  /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
                 84:  /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
                 85:  /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6160	100.0	6160	1	PCT-US97-06830-3	Sequence 3, Appli
	2	6160	100.0	6160	10	US-08-665-401-3	Sequence 3, Appli
	3	6160	100.0	6160	34	US-09-902-432-3	Sequence 3, Appli
	4	5204.2	84.5	5236	1	PCT-US02-16173A-2063	Sequence 2063, Ap
	5	5204.2	84.5	5236	41	US-10-191-803-279	Sequence 279, App
	6	5200	84.4	5200	13	US-08-978-277A-3	Sequence 3, Appli
	7	4831.4	78.4	5134	1	PCT-US97-06830-1	Sequence 1, Appli
	8	4831.4	78.4	5134	10	US-08-665-401-1	Sequence 1, Appli
	9	4831.4	78.4	5134	34	US-09-902-432-1	Sequence 1, Appli
	10	4701.4	76.3	5074	13	US-08-978-277A-1	Sequence 1, Appli
	11	4548.6	73.8	6199	80	US-60-360-207-12210	Sequence 12210, A
	12	2181.2	35.4	7651	30	US-09-770-174-4018	Sequence 4018, Ap
	13	2178	35.4	6603	71	US-60-278-232-10143	Sequence 10143, A
c	14	2178	35.4	6607	68	US-60-245-228-517	Sequence 517, App
	15	2172.8	35.3	6632	76	US-60-324-185-30580	Sequence 30580, A
	16	2160.4	35.1	8660	25	US-09-644-867-6364	Sequence 6364, Ap
	17	2160.4	35.1	8660	25	US-09-644-869-9117	Sequence 9117, Ap
	18	2160.4	35.1	8660	25	US-09-652-911-10117	Sequence 10117, A
	19	2160.4	35.1	8660	25	US-09-652-913-10715	Sequence 10715, A
	20	2160.4	35.1	8660	25	US-09-652-918-8758	Sequence 8758, Ap
	21	2160.4	35.1	8660	26	US-09-667-760-2000	Sequence 2000, Ap

22	2160.4	35.1	8660	27	US-09-699-997-11667	Sequence 11667, A
23	2160.4	35.1	8660	27	US-09-699-999-6654	Sequence 6654, Ap
24	2160.4	35.1	8660	28	US-09-700-000-5679	Sequence 5679, Ap
25	2160.4	35.1	8660	28	US-09-710-280-2315	Sequence 2315, Ap
26	2160.4	35.1	8660	28	US-09-710-286-3428	Sequence 3428, Ap
27	2160.4	35.1	8660	28	US-09-716-472-4599	Sequence 4599, Ap
28	2160.4	35.1	8660	28	US-09-716-953-2245	Sequence 2245, Ap
29	2160.4	35.1	8660	29	US-09-726-789-3701	Sequence 3701, Ap
30	2160.4	35.1	8660	29	US-09-726-802-2407	Sequence 2407, Ap
31	2160.4	35.1	8660	29	US-09-726-805-1797	Sequence 1797, Ap
32	2160.4	35.1	8660	29	US-09-726-806-5392	Sequence 5392, Ap
33	2159.6	35.1	8492	23	US-09-606-776-5120	Sequence 5120, Ap
34	2144.4	34.8	6608	16	US-09-220-132-58	Sequence 58, Appl
35	2144.4	34.8	6608	29	US-09-738-877-1	Sequence 1, Appli
36	2144.4	34.8	6608	33	US-09-880-107-3439	Sequence 3439, Ap
37	2144.4	34.8	6608	41	US-10-155-653-58	Sequence 58, Appl
38	2108.8	34.2	8886	76	US-60-324-185-30584	Sequence 30584, A
39	2106.2	34.2	8869	71	US-60-278-232-10145	Sequence 10145, A
40	2106.2	34.2	8888	65	US-60-213-359-7019	Sequence 7019, Ap
41	2094.6	34.0	8877	61	US-60-172-360-24882	Sequence 24882, A
42	2055.6	33.4	8243	17	US-09-397-022-4579	Sequence 4579, Ap
43	2055.6	33.4	8243	18	US-09-404-549-4757	Sequence 4757, Ap
44	2055.6	33.4	8243	18	US-09-404-549A-4757	Sequence 4757, Ap
45	2055.6	33.4	8243	18	US-09-428-151A-10174	Sequence 10174, A

ALIGNMENTS

```
RESULT 1
PCT-US97-06830-3
; Sequence 3, Application PC/TUS9706830
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/06830
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6160 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
PCT-US97-06830-3

  Query Match             100.0%;  Score 6160;  DB 1;  Length 6160;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 6160;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy        361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600

Qy        601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660

Qy        661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy        721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy        781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy        841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy        901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy        961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy   1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy   1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy   1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy   1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy   1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy   1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy   1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy   1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy   1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy   1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy   1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy   1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy   1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy   1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy   1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy   1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy   1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy   2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy   2101 GGCCAAGGACACTGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GGCCAAGGACACTGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy   2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy   2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy   2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy   2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400

Qy   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460

Qy   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520

Qy   2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580

Qy   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760

Qy   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120

Qy   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
```

```
Qy     3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy     3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy     3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy     3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy     3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy     3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy     3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy     3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy     3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy     3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy     3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy     3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy     3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy     3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy     4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy     4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy     4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy     4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy     4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy     4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380

Qy     4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy     4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy     4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy     4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy     4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy     4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy     4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy     4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860

Qy     4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920

Qy     4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980

Qy     4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040

Qy     5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100

Qy     5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160

Qy     5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220

Qy     5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280

Qy     5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340

Qy     5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400

Qy    5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460

Qy    5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520

Qy    5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAA 5580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAA 5580

Qy    5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640

Qy    5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700

Qy    5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760

Qy    5761 CAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTC 5820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5761 CAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTC 5820

Qy    5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880

Qy    5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940

Qy    5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000

Qy    6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTC 6060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTC 6060

Qy    6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120

Qy    6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
           ||||||||||||||||||||||||||||||||||||||||
Db    6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
```

```
RESULT 2
US-08-665-401-3
; Sequence 3, Application US/08665401
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/665,401
;      FILING DATE:  18-JUN-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6160 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-665-401-3

  Query Match            100.0%;  Score 6160;  DB 10;  Length 6160;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 6160;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy        361 CCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
```

```
Qy     601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660

Qy     661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy     721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy     781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy     841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy     901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy     961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy    1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy    1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy    1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy    1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy    1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy    1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy    1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy    1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy    1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy    1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy    1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy    1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy    1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy    1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy    1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy    1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy    1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy    2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy    2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy    2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy    2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy    2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy    2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400

Qy    2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460

Qy    2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520

Qy    2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580

Qy    2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy    2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy    2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760

Qy    2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120

Qy   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180

Qy   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380

Qy   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy   4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy   4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy   4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy   4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860

Qy   4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920

Qy   4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
```

```
Qy   4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040

Qy   5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100

Qy   5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160

Qy   5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220

Qy   5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280

Qy   5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340

Qy   5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400

Qy   5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460

Qy   5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTATGTAGTCCTCC 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTATGTAGTCCTCC 5520

Qy   5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTGTATTATGCAA 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTGTATTATGCAA 5580

Qy   5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640

Qy   5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700

Qy   5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760

Qy   5761 CAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTC 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 CAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTC 5820

Qy   5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880

Qy   5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940

Qy   5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000

Qy   6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTC 6060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTC 6060

Qy   6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120

Qy   6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
          ||||||||||||||||||||||||||||||||||||||||
Db   6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
```

```
RESULT 3
US-09-902-432-3
; Sequence 3, Application US/09902432
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 6160
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-902-432-3

  Query Match             100.0%;  Score 6160;  DB 34;  Length 6160;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 6160;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy     61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy    121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy    181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy    241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy    301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy    361 CCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAAC 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy    421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy    481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy    541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600

Qy      601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660

Qy      661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy      721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy      781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy      841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy      901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy      961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy     1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy     1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy     1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy     1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy     1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy     1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy     1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy     1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy     1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy     1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy     1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy     1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy     1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy     1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy     1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy     1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy     1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy     2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy     2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy     2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy     2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy     2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy     2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400

Qy     2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460

Qy     2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520

Qy     2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580

Qy     2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy     2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy     2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
```

```
Qy   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120

Qy   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180

Qy   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380

Qy   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy   4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy   4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy   4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy   4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860

Qy   4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920

Qy   4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980

Qy     4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040

Qy     5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100

Qy     5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160

Qy     5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220

Qy     5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280

Qy     5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340

Qy     5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400

Qy     5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460

Qy     5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520

Qy     5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTGTATTATGCAA 5580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTGTATTATGCAA 5580

Qy     5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640

Qy     5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700

Qy     5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760

Qy     5761 CAGAAATAAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTC 5820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5761 CAGAAATAAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTC 5820

Qy     5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880

Qy     5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940

Qy     5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000

Qy     6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTC 6060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTC 6060

Qy     6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120

Qy     6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
            ||||||||||||||||||||||||||||||||||||||||
Db     6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160


RESULT 4
PCT-US02-16173A-2063
; Sequence 2063, Application PC/TUS0216173A
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Higgs, Brandon
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5089-WO
;  CURRENT APPLICATION NUMBER: PCT/US02/16173A
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: US 60/292,335
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/297,523
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,925
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,810
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,807
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,808
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/315,047
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/324,928
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/330,867
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/330,462
;  PRIOR FILING DATE: 2001-10-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 2221
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2063
;   LENGTH: 5236
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. NM_057103
PCT-US02-16173A-2063

  Query Match             84.5%;  Score 5204.2;  DB 1;  Length 5236;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 5223;  Conservative    0;  Mismatches   13;  Indels    1;  Gaps    1;

Qy        1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy       61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy      121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy      181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy      241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy      301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy      361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy      421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy      481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy      541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600

Qy      601 GGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660

Qy      661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy      721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy      781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy      841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy      901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy      961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy     1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy     1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy     1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy     1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy     1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy     1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy     1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy     1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy     1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy     1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy     1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy     1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy     1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy     1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy     1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy     1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy     1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy     2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy     2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy     2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy     2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy     2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy     2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
```

```
Qy   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAGAACAAGCCA 2460

Qy   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTGATGTCCCAGCCGTCG 2520

Qy   2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |||
Db   2521 TGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATGCGG 2580

Qy   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760

Qy   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120

Qy   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180

Qy   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380

Qy   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy     4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
            |||||||||||||||||||||||||        |  ||||||||||||||||||||||||
Db     4621 ATGATTCACAGACCCAGGTTCCTGCAATGCAGGCTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy     4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
            ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     4681 GGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy     4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy     4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db     4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGAT-CAAA 4859

Qy     4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4860 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4919

Qy     4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4920 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4979

Qy     4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
            ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4980 ATCGCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5039

Qy     5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5040 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5099

Qy     5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5100 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5159

Qy     5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5160 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5219

Qy     5221 TCCAATGATTTCCACCC 5237
            |||||||||||||||||
Db     5220 TCCAATGATTTCCACCC 5236
```

```
RESULT 5
US-10-191-803-279
; Sequence 279, Application US/10191803
; GENERAL INFORMATION:
;  APPLICANT: MENDRICK, Donna
;  APPLICANT:  PORTER, Mark
;  APPLICANT:  JOHNSON, Kory
;  APPLICANT:  HIGGS, Brandon
;  APPLICANT:  CASTLE, Arthur
;  APPLICANT:  ELASHOFF, Michael
;  TITLE OF INVENTION: Cardiotoxin Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5090US
;  CURRENT APPLICATION NUMBER: US/10/191,803
;  CURRENT FILING DATE:  2002-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,819
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/305,623
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/369,351
;  PRIOR FILING DATE: 2002-04-03
;  PRIOR APPLICATION NUMBER: US 60/377,611
;  PRIOR FILING DATE: 2002-05-06
;  NUMBER OF SEQ ID NOS: 1140
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 279
;  LENGTH: 5236
;  TYPE: DNA
;  ORGANISM: Rattus norvegicus
;  FEATURE:
;  OTHER INFORMATION: Genbank Accession No. NM_057103
US-10-191-803-279

  Query Match             84.5%;  Score 5204.2;  DB 41;  Length 5236;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 5223;  Conservative    0;  Mismatches   13;  Indels    1;  Gaps    1;
```

```
Qy        1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy       61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy      121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy      181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy      241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy      301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy      361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy      421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy      481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy      541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600

Qy      601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660

Qy      661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy      721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy      781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy      841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy      901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy     961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy    1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy    1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy    1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy    1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy    1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy    1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy    1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy    1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy    1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy    1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy    1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy    1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy    1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy    1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy    1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy    1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy    1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy    2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy    2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy    2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy    2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy    2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy    2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400

Qy    2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
           |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db    2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAGAACAAGCCA 2460

Qy    2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
           |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db    2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTGATGTCCCAGCCGTCG 2520

Qy    2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
           ||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |||
Db    2521 TGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATGCGG 2580

Qy    2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy    2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy    2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760

Qy    2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy    2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy    2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy    2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy    3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy    3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
```

```
Qy   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180

Qy   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380

Qy   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy   4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
          |||||||||||||||||||||||||         |    |||||||||||||||||||||
Db   4621 ATGATTCACAGACCCAGGTTCCTGCAATGCAGGCTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy   4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
          ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 GGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy   4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy   4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db   4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGAT-CAAA 4859

Qy   4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4860 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4919

Qy   4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4920 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4979

Qy   4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
          ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4980 ATCGCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5039

Qy   5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5040 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5099

Qy   5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5100 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5159

Qy   5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5160 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5219

Qy   5221 TCCAATGATTTCCACCC 5237
          |||||||||||||||||
Db   5220 TCCAATGATTTCCACCC 5236
```

RESULT 6

```
US-08-978-277A-3
; Sequence 3, Application US/08978277A
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,277A
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5200 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-978-277A-3

  Query Match             84.4%;  Score 5200;  DB 13;  Length 5200;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 5200;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 60

Qy       93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 120

Qy      153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 180

Qy      213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 240

Qy      273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 300

Qy      333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 360

Qy      393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 420

Qy      453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 480

Qy      513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 540

Qy      573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 600

Qy      633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 660

Qy      693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 720

Qy      753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 780

Qy      813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 840

Qy      873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 900

Qy      933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 960

Qy      993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1020

Qy     1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1080

Qy     1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1140

Qy     1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1200

Qy     1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1260

Qy     1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1320

Qy     1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1380

Qy     1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1440
```

```
Qy   1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1500

Qy   1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1560

Qy   1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1620

Qy   1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1680

Qy   1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1740

Qy   1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1800

Qy   1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1860

Qy   1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1920

Qy   1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 1980

Qy   2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2040

Qy   2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2100

Qy   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2160

Qy   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2220

Qy   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2280

Qy   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2340

Qy   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2400

Qy   2433 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2460

Qy   2493 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2520

Qy   2553 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2580

Qy   2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2640

Qy   2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2700

Qy   2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2760

Qy   2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2820

Qy   2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2880

Qy   2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2940

Qy   2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3000

Qy   3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3060

Qy   3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3120

Qy   3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3180

Qy   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3240

Qy   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3300

Qy   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3360

Qy   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3420

Qy   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3480

Qy   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3540

Qy   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3600

Qy   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3601 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3660

Qy   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3720

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3780

Qy   3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3840

Qy   3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3900

Qy   3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3960

Qy   3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4020

Qy   4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4080

Qy   4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4140

Qy   4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4200

Qy   4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4260

Qy   4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4320

Qy   4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4380

Qy   4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4440

Qy   4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4500

Qy   4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4560

Qy   4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4652
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4620

Qy   4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4680

Qy   4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4740

Qy   4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGA 4832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGA 4800

Qy   4833 AGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAA 4892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 AGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAA 4860

Qy   4893 AGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACC 4952
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 AGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACC 4920

Qy   4953 AACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAA 5012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 AACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAA 4980

Qy   5013 ATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCA 5072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 ATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCA 5040

Qy   5073 AAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGA 5132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 AAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGA 5100

Qy   5133 ATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAG 5192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 ATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAG 5160

Qy   5193 AGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTC 5232
          ||||||||||||||||||||||||||||||||||||||||
Db   5161 AGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTC 5200
```

```
RESULT 7
PCT-US97-06830-1
; Sequence 1, Application PC/TUS9706830
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/06830
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5134 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
```

```
;      TOPOLOGY:  linear
;      MOLECULE TYPE:  cDNA
;      HYPOTHETICAL:  NO
;      ANTI-SENSE:  NO
;      FRAGMENT TYPE:
;      ORIGINAL SOURCE:
PCT-US97-06830-1

  Query Match             78.4%;  Score 4831.4;  DB 1;  Length 5134;
  Best Local Similarity   97.9%;  Pred. No. 0;
  Matches 5048;  Conservative    0;  Mismatches   76;  Indels   31;  Gaps   14;

Qy       1016 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 1075
Db          1 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 60

Qy       1076 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 1135
Db         61 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 120

Qy       1136 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 1195
Db        121 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 180

Qy       1196 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 1255
Db        181 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 240

Qy       1256 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 1315
Db        241 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 300

Qy       1316 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 1375
Db        301 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 360

Qy       1376 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 1435
Db        361 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 420

Qy       1436 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 1495
Db        421 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 480

Qy       1496 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 1555
Db        481 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 540

Qy       1556 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 1615
Db        541 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 600

Qy       1616 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 1675
Db        601 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 660

Qy       1676 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 1735
Db        661 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 720

Qy       1736 AGGCCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGA 1795
Db        721 AGGCCCGCTGGAAGCA-CCCAGGATGGGGAAGCTGAGGAAGGAACTACTTC----GTGGA 775

Qy       1796 GAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 1855
Db        776 GAGAAGAAGAGGAAGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 835

Qy       1856 ACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 1915
Db        836 AC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 894

Qy       1916 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 1975
Db        895 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 954

Qy       1976 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 2035
Db        955 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 1014

Qy       2036 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGA 2095
Db       1015 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCA---GATAT 1071

Qy       2096 AGGAGGGCCAAGGACACTGGGA-GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 2154
Db       1072 AAGAGGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 1131

Qy       2155 AAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 2214
Db       1132 AAGAAGCC-GAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 1190

Qy       2215 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 2274
Db       1191 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 1250

Qy       2275 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCG 2334
Db       1251 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGAAG 1310

Qy       2335 AAGACTCTAGTGTAG--------AGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGA 2386
Db       1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy       2387 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA 2446
Db       1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy       2447 -GCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTA 2505
Db       1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy       2506 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAG 2565
Db       1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy       2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
Db       1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGT---AGTGTCCGAGGAGCTCA 1607

Qy       2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
Db       1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy       2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
Db       1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy       2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
Db       1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy       2806 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 2865
Db       1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy       2866 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 2925
Db       1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy       2926 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 2985
Db       1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy       2986 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 3045
```

```
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   3046 AGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 3105
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   3106 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 3165
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   3166 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 3225
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   3226 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 3285
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   3286 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGG 3345
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   3346 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 3405
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   3406 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 3465
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   3466 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 3525
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   3526 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 3585
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   3586 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 3645
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   3646 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCG 3705
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   3706 TTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 3765
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   3766 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 3825
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   3826 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 3885
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   3886 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 3945
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   3946 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 4005
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   4006 AAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 4065
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   4066 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 4125
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   4126 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 4185
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   4186 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 4245
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   4246 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 4305
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   4306 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 4365
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   4366 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 4425
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   4426 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 4485
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   4486 CAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 4545
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   4546 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 4605
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   4606 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGG 4665
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   4666 AGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCT 4785
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGCTCAG--CTCGGAAATGCT 3765

Qy   4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 3825

Qy   4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
Db   3826 ACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 3884

Qy   4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
Db   3885 CAGAGGCA---GTGTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 3941

Qy   4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
Db   3942 TAACCGAGGAGCG-ATGCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 4000

Qy   5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 5085
Db   4001 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 4060

Qy   5086 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 5145
Db   4061 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 4120

Qy   5146 TCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATC 5205
Db   4121 TCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCATGAGATC 4179
```

```
Qy   5206 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 5265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4180 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 4239

Qy   5266 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4240 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4299

Qy   5326 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4300 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4359

Qy   5386 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 5445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4360 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 4419

Qy   5446 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 5505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4420 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 4479

Qy   5506 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 5565
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4480 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 4539

Qy   5566 TTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 5625
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4540 TTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 4599

Qy   5626 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 5685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4600 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 4659

Qy   5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 5745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4660 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 4719

Qy   5746 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATAT 5805
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4720 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATAT 4779

Qy   5806 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 5865
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4780 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 4839

Qy   5866 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 5925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4840 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 4899

Qy   5926 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 5985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4900 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 4959

Qy   5986 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 6045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4960 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 5019

Qy   6046 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 6105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5020 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 5079

Qy   6106 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5080 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 5134
```

```
RESULT 8
US-08-665-401-1
; Sequence 1, Application US/08665401
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/665,401
;      FILING DATE:  18-JUN-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5134 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-665-401-1

  Query Match             78.4%;  Score 4831.4;  DB 10;  Length 5134;
  Best Local Similarity   97.9%;  Pred. No. 0;
  Matches 5048;  Conservative    0;  Mismatches   76;  Indels   31;  Gaps   14;

Qy   1016 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 1075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 60

Qy   1076 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 1135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 120

Qy   1136 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 1195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 180

Qy   1196 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 1255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 240

Qy   1256 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 1315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 300

Qy   1316 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 1375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 360

Qy   1376 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 1435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 420

Qy   1436 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 1495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    421 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 480

Qy   1496 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 1555
Db    481 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 540

Qy   1556 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 1615
Db    541 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 600

Qy   1616 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 1675
Db    601 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 660

Qy   1676 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 1735
Db    661 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 720

Qy   1736 AGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGA 1795
Db    721 AGGGCCGCTGGAAGCA-CCCAGGATGGGGAAGCTGAGGAAGGAACTACTTC----GTGGA 775

Qy   1796 GAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 1855
Db    776 GAGAAGAAGAGGAAGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 835

Qy   1856 ACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 1915
Db    836 AC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 894

Qy   1916 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 1975
Db    895 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 954

Qy   1976 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 2035
Db    955 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 1014

Qy   2036 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGA 2095
Db   1015 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCA---GATAT 1071

Qy   2096 AGGAGGGCCAAGGACACTGGGA-GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 2154
Db   1072 AAGAGGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 1131

Qy   2155 AAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 2214
Db   1132 AAGAAGCC-GAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 1190

Qy   2215 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 2274
Db   1191 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 1250

Qy   2275 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCG 2334
Db   1251 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGAAG 1310

Qy   2335 AAGACTCTAGTGTAG--------AGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGA 2386
Db   1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy   2387 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA 2446
Db   1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy   2447 -GCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTA 2505
Db   1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy   2506 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAG 2565
Db   1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy   2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
Db   1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGT---AGTGTCCGAGGAGCTCA 1607

Qy   2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
Db   1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy   2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
Db   1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy   2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
Db   1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy   2806 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 2865
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy   2866 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 2925
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy   2926 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 2985
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy   2986 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 3045
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   3046 AGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 3105
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   3106 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 3165
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   3166 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 3225
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   3226 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 3285
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   3286 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGG 3345
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   3346 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 3405
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   3406 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 3465
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   3466 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 3525
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   3526 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 3585
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   3586 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 3645
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627
```

```
Qy   3646 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCG 3705
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   3706 TTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 3765
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   3766 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 3825
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   3826 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 3885
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   3886 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 3945
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   3946 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 4005
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   4006 AAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 4065
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   4066 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 4125
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   4126 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 4185
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   4186 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 4245
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   4246 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 4305
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   4306 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 4365
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   4366 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 4425
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   4426 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 4485
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   4486 CAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 4545
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   4546 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 4605
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   4606 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGG 4665
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   4666 AGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCT 4785
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGCTCAG--CTCGGAAATGCT 3765

Qy   4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 3825

Qy   4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
Db   3826 ACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 3884

Qy   4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
Db   3885 CAGAGGCA---GTGTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 3941

Qy   4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
Db   3942 TAACCGAGGAGCG-ATGCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 4000

Qy   5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 5085
Db   4001 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 4060

Qy   5086 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 5145
Db   4061 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 4120

Qy   5146 TCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATC 5205
Db   4121 TCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCATGAGATC 4179

Qy   5206 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 5265
Db   4180 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 4239

Qy   5266 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5325
Db   4240 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4299

Qy   5326 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5385
Db   4300 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4359

Qy   5386 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 5445
Db   4360 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 4419

Qy   5446 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 5505
Db   4420 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 4479

Qy   5506 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 5565
Db   4480 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 4539

Qy   5566 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 5625
Db   4540 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 4599

Qy   5626 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 5685
Db   4600 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 4659

Qy   5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 5745
Db   4660 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 4719

Qy   5746 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 5805
Db   4720 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 4779

Qy   5806 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 5865
```

```
Db   4780 TACCATGATTTTTCCCCCCTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 4839

Qy   5866 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 5925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4840 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 4899

Qy   5926 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTGATTGTACTCTTCTATGCT 5985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4900 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTGATTGTACTCTTCTATGCT 4959

Qy   5986 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 6045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4960 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 5019

Qy   6046 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 6105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5020 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 5079

Qy   6106 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5080 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 5134
```

```
RESULT 9
US-09-902-432-1
; Sequence 1, Application US/09902432
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 5134
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-902-432-1

  Query Match            78.4%;  Score 4831.4;  DB 34;  Length 5134;
  Best Local Similarity  97.9%;  Pred. No. 0;
  Matches 5048;  Conservative    0;  Mismatches   76;  Indels   31;  Gaps   14;

Qy   1016 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 1075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 60

Qy   1076 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 1135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 120

Qy   1136 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 1195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 180

Qy   1196 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 1255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 240

Qy   1256 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 1315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 300

Qy   1316 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 1375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 360

Qy   1376 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 1435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 420

Qy   1436 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 1495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 480

Qy   1496 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 1555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 540

Qy   1556 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 1615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 600

Qy   1616 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 1675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 660

Qy   1676 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 1735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 720

Qy   1736 AGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGA 1795
          |||||||||||||||| ||||||||||||||||||||||||||||||||||    |  ||
Db    721 AGGGCCGCTGGAAGCA-CCCAGGATGGGGAAGCTGAGGAAGGAACTACTTC----GTGGA 775

Qy   1796 GAAGAAGAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 1855
          ||  |  |||| |||||||||||||||||||||||||||||||||||||||||||||||
Db    776 GAGAAGAAGAGGAAGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 835

Qy   1856 ACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 1915
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    836 AC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 894

Qy   1916 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 1975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    895 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 954

Qy   1976 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 2035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    955 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 1014

Qy   2036 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGA 2095
          |||||||||||||||||||||||||||||||||||||||||||||||||||     |||
Db   1015 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCA---GATAT 1071

Qy   2096 AGGAGGGCCAAGGACACTGGGA-GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 2154
          | |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db   1072 AAGAGGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 1131

Qy   2155 AAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 2214
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1132 AAGAAGCC-GAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 1190

Qy   2215 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 2274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 1250

Qy   2275 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCG 2334
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db   1251 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGAAG 1310

Qy   2335 AAGACTCTAGTGTAG--------AGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGA 2386
             |    | | |||          |||||||||||||||||||| |||||||||
Db   1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy   2387 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA 2446
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430
```

```
Qy   2447 -GCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTA 2505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||| 
Db   1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy   2506 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAG 2565
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy   2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
          ||||||||||| |||||||||         |||   |||     ||||||||||||||
Db   1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGT---AGTGTCCGAGGAGCTCA 1607

Qy   2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy   2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy   2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy   2806 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 2865
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy   2866 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 2925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy   2926 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 2985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy   2986 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 3045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   3046 AGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 3105
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   3106 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 3165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   3166 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 3225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   3226 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 3285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   3286 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGG 3345
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   3346 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 3405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   3406 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 3465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   3466 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 3525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   3526 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 3585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   3586 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 3645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   3646 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCG 3705
          |||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   3706 TTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 3765
          |   |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   3766 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 3825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   3826 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 3885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   3886 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 3945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   3946 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 4005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   4006 AAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 4065
          || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   4066 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 4125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   4126 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 4185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   4186 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 4245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   4246 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 4305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   4306 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 4365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   4366 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 4425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   4426 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 4485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   4486 CAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 4545
          |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   4546 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 4605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   4606 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGG 4665
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||     |  | ||||||||||
Db     3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy     4666 AGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy     4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGCTCAGCCTCGGAAATGCT 4785
            ||||||||||||||||||||||||||||||||||||||||||||||  |||||||||||
Db     3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGCTCAG--CTCGGAAATGCT 3765

Qy     4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 3825

Qy     4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
            ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3826 ACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 3884

Qy     4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
            |||||||    |   |||||||||||||||||||||||||||||||||||||||||||||
Db     3885 CAGAGGCA---GTGTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 3941

Qy     4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
            ||||||||||| |    |||||||||||||||||||||||||||||||||||||||||||
Db     3942 TAACCGAGGAGCG-ATGCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 4000

Qy     5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 5085
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4001 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 4060

Qy     5086 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 5145
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4061 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 4120

Qy     5146 TCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATC 5205
            |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db     4121 TCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCATGAGATC 4179

Qy     5206 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 5265
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4180 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 4239

Qy     5266 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5325
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4240 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4299

Qy     5326 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5385
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4300 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4359

Qy     5386 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 5445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4360 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 4419

Qy     5446 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 5505
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4420 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 4479

Qy     5506 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTCTAACGTTTAAGCACATGCT 5565
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4480 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTCTAACGTTTAAGCACATGCT 4539

Qy     5566 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 5625
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4540 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 4599

Qy     5626 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 5685
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4600 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 4659

Qy     5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 5745
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4660 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 4719

Qy     5746 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 5805
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4720 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 4779

Qy     5806 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 5865
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4780 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 4839

Qy     5866 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 5925
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4840 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 4899

Qy     5926 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 5985
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4900 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 4959

Qy     5986 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 6045
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4960 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 5019

Qy     6046 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 6105
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5020 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 5079

Qy     6106 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5080 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 5134
```

RESULT 10
US-08-978-277A-1

```
; Sequence 1, Application US/08978277A
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,277A
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5074 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
```

```
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-978-277A-1

  Query Match             76.3%;  Score 4701.4;  DB 13;  Length 5074;
  Best Local Similarity   96.8%;  Pred. No. 0;
  Matches 4988;  Conservative    0;  Mismatches   76;  Indels   91;  Gaps   15;

Qy   1016 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 1075
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 60

Qy   1076 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 1135
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 120

Qy   1136 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 1195
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 180

Qy   1196 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 1255
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 240

Qy   1256 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 1315
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 300

Qy   1316 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 1375
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 360

Qy   1376 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 1435
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 420

Qy   1436 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 1495
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 480

Qy   1496 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 1555
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 540

Qy   1556 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 1615
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 600

Qy   1616 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 1675
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 660

Qy   1676 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 1735
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 720

Qy   1736 AGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGA 1795
          |||||||||||||||| ||||||||||||||||||||||||||||||||||    |  ||
Db    721 AGGGCCGCTGGAAGCA-CCCAGGATGGGGAAGCTGAGGAAGGAACTACTTC----GTGGA 775

Qy   1796 GAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 1855
          ||  |  |||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db    776 GAGAAGAAGAGGAAGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 835

Qy   1856 ACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 1915
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    836 AC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 894

Qy   1916 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 1975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    895 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 954

Qy   1976 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 2035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    955 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 1014

Qy   2036 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGA 2095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db   1015 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCA---GATAT 1071

Qy   2096 AGGAGGGCCAAGGACACTGGGA-GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 2154
          | |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1072 AAGAGGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 1131

Qy   2155 AAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 2214
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1132 AAGAAGCC-GAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 1190

Qy   2215 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 2274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 1250

Qy   2275 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCG 2334
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db   1251 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGAAG 1310

Qy   2335 AAGACTCTAGTGTAGAGCA--------GTTGTCCACTGAGATCGAACCGAGTAGAGAAGA 2386
            |     | | |||            |||||||||||||||||||||| ||||||||||
Db   1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy   2387 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA 2446
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy   2447 -GCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTA 2505
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy   2506 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAG 2565
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy   2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
          ||||||||||| |||||||||||        |||          ||||||||||||||||
Db   1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy   2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy   2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy   2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy   2806 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 2865
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy   2866 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 2925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy   2926 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 2985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy   2986 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 3045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   3046 AGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 3105
```

```
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   3106 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 3165
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   3166 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 3225
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   3226 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 3285
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   3286 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGG 3345
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   3346 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 3405
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   3406 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 3465
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   3466 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 3525
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   3526 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 3585
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   3586 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 3645
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   3646 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCG 3705
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   3706 TTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 3765
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   3766 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 3825
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   3826 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 3885
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   3886 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 3945
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   3946 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 4005
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   4006 AAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 4065
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   4066 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 4125
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   4126 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 4185
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   4186 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 4245
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   4246 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 4305
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   4306 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 4365
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   4366 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 4425
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   4426 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 4485
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   4486 CAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 4545
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   4546 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 4605
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   4606 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGG 4665
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   4666 AGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCT 4785
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGG--CTCAGCTCGGAAATGCT 3765

Qy   4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 3825

Qy   4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
Db   3826 ACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 3884

Qy   4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
Db   3885 CAGAGGC---AGTGTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 3941

Qy   4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
Db   3942 TAACCGAGGAGCG-ATGCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 4000

Qy   5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 5085
Db   4001 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 4060

Qy   5086 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 5145
Db   4061 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 4120

Qy   5146 TCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATC 5205
Db   4121 TCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCATGAGATC 4179

Qy   5206 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 5265
Db   4180 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 4239
```

```
Qy     5266 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5325
            |||||||||||||||||||
Db     4240 TCATCGGGAGCTAGAGCCAG---------------------------------------- 4259

Qy     5326 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5385
                                |||||||||||||||||||||||||||||||||||||||||
Db     4260 --------------------CTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4299

Qy     5386 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 5445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4300 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 4359

Qy     5446 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 5505
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4360 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 4419

Qy     5506 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 5565
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4420 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 4479

Qy     5566 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 5625
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4480 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 4539

Qy     5626 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 5685
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4540 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 4599

Qy     5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 5745
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4600 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 4659

Qy     5746 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 5805
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4660 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 4719

Qy     5806 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 5865
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4720 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 4779

Qy     5866 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 5925
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4780 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 4839

Qy     5926 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 5985
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4840 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 4899

Qy     5986 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 6045
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4900 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 4959

Qy     6046 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 6105
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4960 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 5019

Qy     6106 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5020 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 5074


RESULT 11
US-60-360-207-12210
; Sequence 12210, Application US/60360207
; GENERAL INFORMATION:
;  APPLICANT: VENTER, J. Craig
;  TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
;  FILE REFERENCE: CL001321
;  CURRENT APPLICATION NUMBER: US/60/360,207
;  CURRENT FILING DATE:  2002-03-01
;  NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 12210
;  LENGTH: 6199
;  TYPE: DNA
;  ORGANISM: HUMAN
US-60-360-207-12210

  Query Match             73.8%;  Score 4548.6;  DB 80;  Length 6199;
  Best Local Similarity   86.6%;  Pred. No. 0;
  Matches 5375;  Conservative    0;  Mismatches  644;  Indels  190;  Gaps   25;

Qy        4 GCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGGA 63
            |||||| |||||| | ||||||||||||||| |||||||||||||||||||||||||||
Db      113 GCCGGCGAGAGTAGGAAAGCCACTGAGCCATGGGTGCAGGCAGTTCCACCGAGCAGCGGA 172

Qy       64 GCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATGGGC 123
            |||||||||||||||| |||||||||||||||||||||| ||||||||||||||||||
Db      173 GCCCCGAGCAGCCGGCGGAGAGCGACACGCCGAGCGAGCTGGAGCTCAGTGGCCATGGGC 232

Qy      124 CCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCAAGC 183
            |||||| ||||| |||||||||||||||| ||||| |||||||||||||||||||||
Db      233 CCGCAGCGGAAGCGTCGGGAGCAGCTGGAGATCCCGCTGACGCGGACCCCGCCACCAAGC 292

Qy      184 TCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATGTCC 243
            ||||||||||||||| ||||||||| | ||||| || ||||||||||||| |||||||
Db      293 TCCCACAGAAGAATGGTCAGCTGTCTGCCGTCAATGGTGTAGCTGAACAAGAAGATGTCC 352

Qy      244 ATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGAC 303
             | |||||||||||| |||||| ||||| ||||||||||||   || ||| |||||||||
Db      353 ACGTCCAAGAGGAAAGCCAGGATGGGCAAGAGGAAGAAGTCACTGTTGAAGATGTTGGAC 412

Qy      304 AGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCA 363
            || |||||||||||||||||| ||||||||||||| | |||||||||| ||||   ||||
Db      413 AGAGAGAGTCAGAAGATGTGAAAGAAAAAGACCGAGCTAAAGAAATGGCAGCCAGTTCCA 472

Qy      364 CAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGA 423
            ||| ||||||||||||||||||||| | ||||||||| ||| | |||||||| |||||||
Db      473 CAGTTGTTGAAGATATCACAAAGGACGAGCAGGAGGAAACACCGGAAATAATCGAACAGA 532

Qy      424 TCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATG 483
            ||||||||||||| | ||||||||||||||||| |||  |||||||||||||| ||||||
Db      533 TCCCTGCTTCAGAGAGCAATGTGGAAGAAATGGCGCAGGCTGCTGAGTCCCAAGCTAATG 592

Qy      484 ATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATA 543
             | || ||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      593 ACGTCGGCTTCAAGAAGGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATA 652

Qy      544 AAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGG 603
            |||| ||||||||||||| ||||| |||||||||||||||||||||||||||||||||||
Db      653 AAAACGAAAAGTCAGATACCGTCCAGCTACTCACTGTCAAGAAGGATGAAGGCGAAGGGG 712

Qy      604 CAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAG 663
            ||||||||| |||||||| ||||||||||| |||||| | |||||||    ||||||| |
Db      713 CAGAAGCCTCCGTCGGAGCAGGAGACCACCAAGAGCCCGGAGTGGAGA---CCGTCGGCG 769

Qy      664 AGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGA 723
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      770 AATCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGA 829

Qy      724 AGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGG 783
            ||||| |||||||||||||||||| |||||||||||||||||||| ||||| | | ||||
Db      830 AGCAAGCACAGAGCAGCACAGAAATTCCCCTTCAAGCCGAATCTGGTCAAGGGACCGAGG 889

Qy      784 AAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAAT 843
            ||||||      | || |||||||||| | ||||||||| || |||||| ||| |||||
Db      890 AAGAAGCAGCCAAAGATGGAGAAGAAAACCGAGAGAAAGAACCTACCAAGCCCCTAGAAT 949

Qy      844 CCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACG 903
            | |||| |||||| |||| || |||||||||||||||||||||||| ||||||||||||
Db      950 CTCCGACCAGCCCTGTCAGCAATGAGACAACATCTTCCTTCAAGAAATTCTTCACTCACG 1009

Qy      904 GTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAA 963
            | ||||||||||||||||||||||||||||||||||| |||| || |||||||||||||
Db     1010 GCTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAACCAAAGGAAGATGATCTGGAAA 1069
```

```
Qy    964 CTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGA 1023
Db   1070 CTTCCGAGAAGAGAAAGGAGCAAGAGGCTGAAAAAGTAGACGAGGAAGAAGGGGAAAAGA 1129

Qy   1024 CAGAGCCAGCC---TCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
Db   1130 CAGAGCCAGCCCCAGCCGAGGAGCAGGAGCCTGCAGAAGGCACAGACCAGGCCAGGTTGT 1189

Qy   1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
Db   1190 CAGCCGACTATGAGAAGGTGGAGTTGCCTTTGGAAGACCAGGTCGGTGACCTGGAGGCAT 1249

Qy   1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
Db   1250 TGTC---GGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGACGGAAGCCC 1306

Qy   1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
Db   1307 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGATGACGAAAGGGCA-- 1364

Qy   1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
Db   1365 -AGGAGGAGCAGAGGTGGAAGGGGATGTGGTGGTGGAAGGATCGGGAGAATCCTTGCCCC 1423

Qy   1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
Db   1424 CTGAGAAACTGGCTGAGACCCAGGAGGTCCCCCAGGAAGCTGAGCCTGTGGAGGAGCTGA 1483

Qy   1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
Db   1484 TGAAGACCAAAGAAGTATGCGTCTCTGGGGGTGACCATACTCAGCTGACAGATCTAAGTC 1543

Qy   1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
Db   1544 CTGAAGAGAAGATGCTACCCAAACACCCCGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1603

Qy   1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
Db   1604 CCTCTCAGGAGAGAATCAAGGTACAGGGAAGTCCCCTGAAGAAGCTCTTCAGCAGTTCGG 1663

Qy   1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAA---ACGAGGAGGTGGGGGAGACG 1617
Db   1664 GCTTAAAGAAGCTCTCCGGGAAGAAGCAGAAGGGGAAGAGAGGAGGAGGCGGGGGAGATG 1723

Qy   1618 AAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGA 1677
Db   1724 AAGAGCCAGGAGAATACCAACACATTCAAACCGAGTCCCCAGAGAGTGCTGACGAGCAGA 1783

Qy   1678 AGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAG 1737
Db   1784 AGGGAGAGAGCTCTGCCTCTTCCCCTGAAGAGCCCGAGGAGATCGCGTGTCTGGAGAAGG 1843

Qy   1738 GGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGA 1797
Db   1844 GGCCATCGGAAGCACCCCAGGAAGCGGAAGCTGAGGAAGGAGCGACTTCCGACGGAGAGA 1903

Qy   1798 AGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAAC 1857
Db   1904 AGAAAAGGGAAGGGATCACCCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAAC 1963

Qy   1858 GGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCA 1917
Db   1964 GGGTCCGAAGACCTTCTGAGAGCGACAAGGAAGAAGAGCTGGATAAGGTCAAGAGTGCCA 2023

Qy   1918 CCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTG 1977
Db   2024 CCTTGTCCTCCACGGAGAGCACGGCGTCTGGAATGCAGGATGAGGTCAGAGCGGTTGGCG 2083

Qy   1978 AGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCAC 2037
Db   2084 AGGAGCAAAGGTCAGAGGAGCCAAAGCGCAGGGTGGATACTTCAGTGTCTTGGGAGGCGT 2143

Qy   2038 TGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAG 2097
Db   2144 TGATTTGTGTCGGATCGTCCAAGAAGAGAGCGAGGAAGGCATCCTCTTCAGATGATGAAG 2203

Qy   2098 GAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAG 2157
Db   2204 GAGGGCCAAGAACACTGGGAGGGGATGGCCACAGAGCGGAGGAGGCTAGCAAAGACAAAG 2263

Qy   2158 AAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTT 2217
Db   2264 AAG------CAGATGCTCTTCCTGCCAGCACCCAGGAACAAGACCAAGCGCACGGAAGTT 2317

Qy   2218 CCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCAT 2277
Db   2318 CCTCACCCGAGCCAGCTGGAAGCCCTTCTGAAGGGGAGGGCGTCTCCACCTGGGAGTCAT 2377

Qy   2278 TTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAG 2337
Db   2378 TTAAGAGATTAGTCACTCCACGAAAAAAATCCAAGTCAAAACTGGAAGAGAGAGCCGAAG 2437

Qy   2338 ACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTT 2397
Db   2438 ACTCCGGTGCAGAGCAGTTGGCCTCCGAGATCGAACCAAGTAGAGAGGAATCTTGGGTTT 2497

Qy   2398 CCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAG 2457
Db   2498 CCATTAAGAAATTTATTCCTGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAGAACAGG 2557

Qy   2458 CCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCG 2517
Db   2558 CCGCCGTTGAAGACTCGGGGCCAGGAGAGATCAATGAGGACGACCCCGACGTCCCAGCTG 2617

Qy   2518 TCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGA---------AGCCC 2568
Db   2618 TTGTGCCTCTGTCTGAGTACGATGCGGTAGAGAGAGAGAAGCTGGAAGCGCAGCGAGCTC 2677

Qy   2569 AGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTA 2628
Db   2678 AGGAGAACGTGGAGCTGCCCCAGCTGAAGGGGGCTGTGTATGTGTCTGAGGAGCTTAGTA 2737

Qy   2629 AGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTG 2688
Db   2738 AGACTCTGGTTCACACCGTGAGTGTCGCGGTCATTGATGGGACCAGGGCAGTCACCAGTG 2797

Qy   2689 TCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAG 2748
Db   2798 CCGAAGAGCGGTCCCCTTCGTGGATATCTGCTTCCATGACAGAACCTCTTGAGCACGCAG 2857

Qy   2749 CGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAA 2808
Db   2858 AGGGAGTGGCCACACCGCCTGTTGGAGAGGTCACTGAAAAAGACATCACTGCAGAAGCAA 2917

Qy   2809 CTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCA 2868
Db   2918 CTCCTGCACTCGCCCAGACTTTACCAGGGGGCAAAGATGCCCATGACGACATAGTCACCA 2977

Qy   2869 GTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTA 2928
Db   2978 GTGAAGTGGATTTTACCTCAGAAGCAGTGACAGCCGCAGAAACCACAGAGGCGCTCCGCG 3037

Qy   2929 CTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTT 2988
Db   3038 CTGAAGAACTTACCGAAGCATCAGGGGCAGAAGAGACCACAGACATGGTGTCTGCAGTTT 3097

Qy   2989 CCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGA 3048
Db   3098 CCCAGCTGTCCGACTCCCCGGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGG 3157

Qy   3049 GTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTG 3108
Db   3158 GTGGCATGCTAGATACGGAAGAACAGGAGCGCCAGACGCAGGCCGTCCTCCAAGCCGTTG 3217

Qy   3109 CAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGT 3168
```

```
Db   3218 CAGACAAAGTGAAAGAGGACTCCCAGGTGCCTGCAACCCAGACTCTGCAGAGAGCAGGGC 3277

Qy   3169 CAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGA 3228
Db   3278 CGAAAGCACTGGAGAAGGTGGAGGAGGTAGAGGAGGACTCCGAGGTGCTGGCTACCGAGA 3337

Qy   3229 AAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCAC 3288
Db   3338 AAGAGAAGGATGTTGTGCCGGAAGGACCCGTGCAGGAAGCTGAAACTGAGCATCTTGCAC 3397

Qy   3289 AGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAG 3348
Db   3398 AGGGCTCCGAGACTGTACAGGCTACCCCAGAGAGCCTTGAAGTTCCTGAAGTCACAGAGG 3457

Qy   3349 ATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCG 3408
Db   3458 ATGTAGACCGTGCCACCACATGCCAGGTTATCAAGCACCAGCAGCTGATGGAACAGGCTG 3517

Qy   3409 TGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAG 3468
Db   3518 TGGCCCCTGAGTCATCTGAAACCTTGACAGACAGTGAGACAAATGGAAGTACTCCCCTCG 3577

Qy   3469 CAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTA 3528
Db   3578 CAGATTCAGACACTCCAAACGGGACACAGCAAGACGAGACCGTTGACAGCCAGGACAGTA 3637

Qy   3529 AAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGA 3588
Db   3638 ATGCCATTGCCGCCGTCAAGCAGTCACAGGTCACTGAAGAGGAGGCAGCTGCTGCTCAGA 3697

Qy   3589 AAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAAC 3648
Db   3698 CGGAGGGGCCTTCAACACCATC---TAGTTTTCCAGCCCAGGAAGAACACAGGGAAAAAC 3754

Qy   3649 CAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTC 3708
Db   3755 CAGGAAGGGATGTTCTAGAACCCACA---CAAGCGCTGGCTGCCGGGGCAGTGCCTATTC 3811

Qy   3709 TGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCA 3768
Db   3812 TGGCAAAGGCTGAGGTGGGTCAAGAGGGTGAGGCTGGCCAGTTTGATGGAGAAAAAGTCA 3871

Qy   3769 AAG--------------AAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAA 3813
Db   3872 AAGACGGACAGTGTGTTAAAGAACTGGAGGTGCCTGTGCACACTGGACCCAACAGTCAAA 3931

Qy   3814 AGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGG 3873
Db   3932 AGACTGCTGACTTGACACGTGACAGTGAAGTAATGGAAGTGGCCAGATGTCAGGAAACTG 3991

Qy   3874 AGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAA 3933
Db   3992 AGAGTAATGAAGAACAGAGTATTAGCCCGGAGAAAAGAGAGATGGAACCGACGTTGAAA 4051

Qy   3934 AGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCG 3993
Db   4052 AGGAGGAAACAGAGACCAAGACAGAGCAAGCCAGTGAAGAACATGAGCAGGAAACAGCTG 4111

Qy   3994 CTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAG 4053
Db   4112 CTCCTGAGCATGAAGGAACCCACCCTAAGCCAGTCCTGACAGCTGACATGCCTCACTCAG 4171

Qy   4054 AGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAG 4113
Db   4172 AGAGGGGAAAGGCACTGGGCAGCCTTGAAGGAAGCCCTTCTCTCCCAGACCAAGACAAAG 4231

Qy   4114 CAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAG 4173
Db   4232 CAGATTGCATAGAGGTTCAAGTTCAAAGCTCAGACACACCAGTCACTCAAACAACCGAAG 4291

Qy   4174 CTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTG 4233
Db   4292 CTGTGAAAAAGGTCGAAGAAACTGTGGCAACTTCAGAGATGGATGAAAGTTTGGAGTGTG 4351

Qy   4234 TAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTC 4293
Db   4352 CAGGTGCGCAATCATTACCAGCTGAGAAGCTCTCCGAAACCGGTGGCTACGGGACTCTTC 4411

Qy   4294 AGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCA 4353
Db   4412 AGCATGGAGAGGACACCGTGCCCCAGGGGCCTGAGTCTCAGGCAGAGTCCATCCCAATAA 4471

Qy   4354 TAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCAT 4413
Db   4472 TAGTAACTCCTGCTCCTGAAAGCATCCTACATTCTGACCTTCAAAGAGAAGTGAGCGCAT 4531

Qy   4414 CCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCA 4473
Db   4532 CCCAGAAACAGAGATCAGATGAAGATAACAAGCCAGATGCTGGTCCTGATGCTGCCGGCA 4591

Qy   4474 AGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGA 4533
Db   4592 AGGAGAGTGCAGCAAGAGAGAAAATCCTCAGGGCTGAACCTGAGATCTTGGAACTTGAGA 4651

Qy   4534 GTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTA 4593
Db   4652 GTAAGAGCAATAAGATTGTCCAGAGTGTCATCCAGACAGCCGTCGACCAGTTTGCACGTA 4711

Qy   4594 CAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTG-CATGCAGG 4652
Db   4712 CAGAAACAGCCCCCGAAACCCACGCTTCTGATTTACAGAATCAGGTTCCTGTGATGCAGG 4771

Qy   4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAG--------------AT 4697
Db   4772 CTGACAGCCAGGGAGCACAACAGATGCTGGACAAAGATGAAAGCGACCTTCAAGTCTCCC 4831

Qy   4698 GCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTG 4757
Db   4832 CCCAAGATGGAACACTCAGTGCCGTAGCCCAGGAAGGACTTGCGGTTTCTGATAGTTCTG 4891

Qy   4758 GAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCA 4817
Db   4892 AAGGCATGAGCAAGGCTTCAGAAATGATCACCACGCTTGCAGTTGAAAGTGCCAGTGTCA 4951

Qy   4818 AAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGAT 4877
Db   4952 AAGAAAGTGTTGAGAAGCTGCCTCTTCAGTGCAAAGAT-GAAAAGGAACATGCTGCTGAC 5010

Qy   4878 GGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAA 4937
Db   5011 GGCCCCCAGCACCAAAGCTTAGCCAAGGCAGAGGCGGATGCCTCTGGAAATCTAACCAAA 5070

Qy   4938 GAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAG 4997
Db   5071 GAGTCCCCAGACACCAACGGACCAAAGCTAACCGAGGAAGGAGATGCCCTGAAAGAAGAA 5130

Qy   4998 GTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAG 5057
Db   5131 AT-----------------------------------GAACAAGGCCCAGACAGAAGAG 5154

Qy   5058 ---GACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCATTG 5114
Db   5155 GACGACCTACAGGAGCCAAAGGGAGACCTGACAGAATCCTAAGACGTGAGTTGCTCAT-- 5212

Qy   5115 TACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACC 5174
Db   5213 ------TGTAAAACTAGAATGTGAAGTGAAGTCACAGAACAAGATGCTGCTGTTGGGACC 5266

Qy   5175 TTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCCA 5234
Db   5267 TTGAGACCAAAATTTCAGAGCCCTTGAGGTGCAGAGAGCAGAGCCGTCCAATGATTTCAA 5326

Qy   5235 CCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATTT 5294
Db   5327 CCCTGCAGAGCACCCCGACAATCCTGAGGCTGCATCGGGAGCTAGAGCCAGCTAACATTT 5386
```

```
Qy   5295 CCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATTT 5354
Db   5387 CCTCTTTTCAAGACTGCCCTTGATTTGCCCCTTGGTG---------------------- 5423

Qy   5355 CCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTAA 5414
Db   5424 -----------------------------------------CCATGTATTTCAGAGTTAA 5442

Qy   5415 GGTCCTGCGTTCTCAACCTGGAACCAAT-TCTGCCATACCTAGTTCCACTTCTCAAACTG 5473
Db   5443 GGTCCTGCTTTCTCCACCTGGAACCAATGTCGGCAATACCTAGTCCCACTTCTCAAACTG 5502

Qy   5474 GAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATT 5533
Db   5503 GAGCCTCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATT 5562

Qy   5534 GTATA-TTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACGGGT 5592
Db   5563 GTATATTTTTTTTCTAAGGTTTAAGCACATGCCTTTTGTATTATGCAGTATATAATGGGT 5622

Qy   5593 GTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAG 5652
Db   5623 GTGCAGCCATAGCGAAGCTTTGAGAAGCCCCAAGCCTCAACTGTAACCTGCAGCAAACAG 5682

Qy   5653 ----ATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGAT 5708
Db   5683 AGAAATAACATTCCTGGCAGGAAGATACAAGTC-TTTTTAAAGTTTACTGATGCTTAACT 5741

Qy   5709 CTGTGGGCTTCTAGTCCTCTGAAAGTGGTGTTTTCCTATGCACAGCGAGCTCAGAAATA 5768
Db   5742 CTGTGGGCTCGTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGTGAGCTCAGAAATA 5801

Qy   5769 AAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGA-TTTTTTCCCCCCTT 5827
Db   5802 AAAACTCCATTTTGAGACATTCAGAATGTCCCAACATTACGACAACATTTTTTTTTTTTT 5861

Qy   5828 TTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTC 5887
Db   5862 TTTTCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTTAGTGTCAGATTAAGCCCCATCTC 5921

Qy   5888 TTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGA 5947
Db   5922 TTAACAATATGGACAGATGAGTGTGCC-ATGGCCATGAGCTATTTCCTAACGCAGAAGGA 5980

Qy   5948 ATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATC 6007
Db   5981 ATTTGTTGT-TACTTTTTGGATTGTACTCTTCTAGGCTGGACCGAATTCATATGCAGATT 6039

Qy   6008 GAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTA 6067
Db   6040 GAAGTGAGACCTGCTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTATTA 6099

Qy   6068 TATCTGTGCCCCTTC-TTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGAT 6126
Db   6100 TATCTGTGCCCCTTCTTTTAAGAACAATGTTACATTATGTTCCTTTGGATAAATTGTGAT 6159

Qy   6127 TTGACAACTGATTTAAATAAACATATTTG 6155
Db   6160 TTGACAACTGATTTCAATAAAAATATTTG 6188
```

RESULT 12
US-09-770-174-4018
; Sequence 4018, Application US/09770174
; GENERAL INFORMATION:
; APPLICANT: Shyjan, Andrew W.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.2057-001
; CURRENT APPLICATION NUMBER: US/09/770,174
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 60/178,835
; PRIOR FILING DATE: 2000-01-28
; NUMBER OF SEQ ID NOS: 4376
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4018
; LENGTH: 7651
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-770-174-4018

```
  Query Match             35.4%;  Score 2181.2;  DB 30;  Length 7651;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 4224;  Conservative    0;  Mismatches 1793;  Indels  537;  Gaps   43;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db    153 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 212

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    213 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 272

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db    273 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 332

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    333 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 392

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db    393 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 452

Qy    269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    453 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 512

Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    513 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 572

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    573 GTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCT 632

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    633 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 692

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    693 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 752

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    753 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 812

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    813 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 857

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    858 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 917

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    918 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCA---GTGGAGGAA 974

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    975 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1034

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
```

```
Db   1035 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1094

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1095 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1154

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1155 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1214

Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1215 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1274

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1275 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1334

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1335 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1394

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1395 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1454

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1455 GAGGAGAGAACCGAAGAGCAGAAA------------------------ACGGAGGTGGAA 1490

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1491 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1550

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1551 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1610

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1611 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1670

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1671 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1730

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1731 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1790

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1791 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1847

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1848 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1907

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1908 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1967

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1968 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2027

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2028 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2087

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2088 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2147

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2148 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2207

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2208 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGG 2267

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2268 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2327

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2328 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2387

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2388 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2447

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2448 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2507

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2508 CTGGAAGAGAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2567

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2568 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2627

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2628 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2687

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2688 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2747

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2748 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAG 2807

Qy   2598 GGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2808 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2867

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2868 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2927

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2928 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2987

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2988 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3047

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3048 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3107

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3108 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTGCT 3167

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3168 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3227
```

```
Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3228 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3287

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3288 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3347

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3348 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3407

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3408 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3467

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3468 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3527

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3528 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3587

Qy   3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3588 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3647

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3648 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3707

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3708 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3767

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3768 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3827

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3828 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3887

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3888 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3947

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3948 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4007

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4008 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4067

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4068 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4127

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4128 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4187

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4188 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4247

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4248 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4307

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4308 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4367

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4368 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4427

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4428 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4487

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4488 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4547

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4548 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4607

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4608 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4667

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4668 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4727

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4728 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGGAGGAT 4781

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4782 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAGCAGTAAACTTGTCCAAAAC 4841

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4842 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4901

Qy   4617 GCTTATGATTCACAGACCCAGGTTC---------------CTGCATGCAGGCTTGACAG 4660
Db   4902 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4961

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4962 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5021

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5022 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5081

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5082 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5141

Qy   4819 ------------------------------------------------------------ 4818
Db   5142 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5201

Qy   4819 ---------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5202 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5261

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5262 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5321

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5322 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5381

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
```

```
Db   5382 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5441

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5442 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5501

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG---------------------CTCATTGTAC 5117
Db   5502 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5561

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5562 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5620

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5621 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5679

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5680 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5738

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5739 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5778

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5779 ------------------------------------------CATATAAATTCTGATTTA 5796

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5797 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5856

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5857 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5913

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTGTATTA 5575
Db   5914 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5973

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5974 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6033

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6034 TATAACCCACGAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6093

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6094 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6153

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6154 ATTACACGGAGTGTGCTAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6213

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6214 ATTGGGAAATTT-------TTCTTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6266

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6267 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6326

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6327 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6386

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6387 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6444

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6445 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6504

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6505 TGCCATTCTTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6564

Qy   6134 CTGATTTAAATAAA 6147
Db   6565 CTGATTTAAATAAA 6578
```

RESULT 13
US-60-278-232-10143
; Sequence 10143, Application US/60278232
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; TITLE OF INVENTION: Polymorphisms Identified Thereby
; FILE REFERENCE: GX-0011 P
; CURRENT APPLICATION NUMBER: US/60/278,232
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 12,557
; SOFTWARE: PERL Program
; SEQ ID NO 10143
;   LENGTH: 6603
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 445786.12
US-60-278-232-10143

```
  Query Match             35.4%;  Score 2178;  DB 71;  Length 6603;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 4222;  Conservative    0;  Mismatches 1795;  Indels  537;  Gaps   43;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db    163 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 222

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    223 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 282

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db    283 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 342

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    343 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 402

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268
Db    403 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 462

Qy    269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    463 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACAGAGGTTGGACAGAGA 522

Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    523 GACTCTGAAGATGTGAGCAAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 582

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
```

```
Db    583 GTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCT 642

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    643 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 702

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    703 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 762

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    763 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 822

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    823 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 867

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    868 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 927

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    928 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCA---GTGGAGGAA 984

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    985 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1044

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1045 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1104

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1105 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1164

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1165 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1224

Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1225 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1284

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1285 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1344

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1345 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1404

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1405 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1464

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1465 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1500

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1501 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAA 1560

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1561 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1620

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1621 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1680

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1681 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1740

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1741 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1800

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1801 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1857

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1858 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1917

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1918 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1977

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1978 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCTTCATTCAAAAAG 2037

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2038 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2097

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2098 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2157

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2158 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2217

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2218 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGG 2277

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2278 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2337

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2338 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2397

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2398 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2457

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2458 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2517

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2518 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2577

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2578 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2637

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2638 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2697

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2698 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2757
```

```
Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2758 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAG 2817

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2818 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2877

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2878 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2937

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2938 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2997

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2998 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3057

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3058 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3117

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3118 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTGCT 3177

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3178 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3237

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3238 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3297

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3298 GAGAGGCGGACTCAAGAGGTCCTCCAGACAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3357

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3358 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3417

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3418 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3477

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3478 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3537

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3538 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3597

Qy   3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3598 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3657

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3658 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3717

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3718 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3777

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3778 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3837

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3838 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3897

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3898 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3957

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3958 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4017

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4018 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4077

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4078 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4137

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4138 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4197

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4198 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4257

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4258 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4317

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4318 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4377

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4378 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4437

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4438 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4497

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4498 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4557

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4558 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4617

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4618 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4677

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4678 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4737

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4738 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGGAGGAT 4791

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4792 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4851

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4852 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4911

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
```

```
Db   4912 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4971

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4972 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5031

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5032 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5091

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5092 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5151

Qy   4819 ------------------------------------------------------------ 4818
Db   5152 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5211

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5212 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5271

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5272 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5331

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5332 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5391

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5392 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5451

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5452 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5511

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5512 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5571

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5572 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5630

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5631 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5689

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5690 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5748

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5749 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5788

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5789 ------------------------------------------CATATAAATTCTGATTTA 5806

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5807 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5866

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5867 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5923

Qy   5533 TGTATATTTTTTTCTAACGTTT----------------AAGCACATGCTTTTTGTATTA 5575
Db   5924 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5983

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5984 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6043

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6044 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6103

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6104 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6163

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6164 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6223

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6224 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6276

Qy   5867 TGTCAGA----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6277 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6336

Qy   5910 GTGCCTAAGGCCATGAGA---------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6337 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6396

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6397 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6454

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6455 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6514

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6515 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6574

Qy   6134 CTGATTTAAATAAA 6147
Db   6575 CTGATTTAAATAAA 6588
```

```
RESULT 14
US-60-245-228-517/c
; Sequence 517, Application US/60245228
; GENERAL INFORMATION:
;  APPLICANT:  Beasley, Ellen
;  TITLE OF INVENTION:  ISOLATED HUMAN CYCLASE PROTEINS, NUCLEIC
;  TITLE OF INVENTION:  ACID MOLECULES ENCODING HUMAN CYCLASE PROTEINS, AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE:  CL000878
;  CURRENT APPLICATION NUMBER: US/60/245,228
;  CURRENT FILING DATE:  2000-11-03
;  NUMBER OF SEQ ID NOS:  630
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 517
;   LENGTH: 6607
;   TYPE: DNA
;   ORGANISM: HUMAN
US-60-245-228-517

  Query Match             35.4%;  Score 2178;  DB 68;  Length 6607;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 4222;  Conservative    0;  Mismatches 1795;  Indels  537;  Gaps   43;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db   6443 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 6384
```

```
Qy      63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    6383 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 6324

Qy     120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db    6323 GGCCCCTGGGCCGAGGCGGCGCCAGACCCCCCCGCGGCCCCAACAATCGCTGCCTGGGAC 6264

Qy     171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    6263 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 6204

Qy     231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268
Db    6203 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 6144

Qy     269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    6143 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 6084

Qy     309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    6083 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 6024

Qy     369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    6023 GTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCT 5964

Qy     429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    5963 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 5904

Qy     489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    5903 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 5844

Qy     549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    5843 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 5784

Qy     609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    5783 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 5739

Qy     669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    5738 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 5679

Qy     729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    5678 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCA---GTGGAGGAA 5622

Qy     789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    5621 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 5562

Qy     849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db    5561 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 5502

Qy     909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db    5501 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 5442

Qy     969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db    5441 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 5382

Qy    1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db    5381 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 5322

Qy    1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db    5321 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 5262

Qy    1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db    5261 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 5202

Qy    1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db    5201 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 5142

Qy    1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db    5141 GAGGAGAGAACCGAAGAGCAGAAA----------------------ACGGAGGTGGAA 5106

Qy    1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db    5105 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAA 5046

Qy    1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db    5045 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 4986

Qy    1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db    4985 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 4926

Qy    1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db    4925 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 4866

Qy    1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db    4865 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 4806

Qy    1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db    4805 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 4749

Qy    1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db    4748 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 4689

Qy    1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db    4688 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 4629

Qy    1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db    4628 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 4569

Qy    1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db    4568 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 4509

Qy    1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db    4508 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 4449

Qy    1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db    4448 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 4389

Qy    2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db    4388 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGG 4329

Qy    2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db    4328 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 4269

Qy    2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
```

```
Db   4268 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 4209

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   4208 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 4149

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   4148 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 4089

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   4088 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 4029

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   4028 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 3969

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   3968 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 3909

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   3908 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 3849

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   3848 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAG 3789

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   3788 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 3729

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   3728 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 3669

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   3668 GCTTCAGTGACAGGACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 3609

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   3608 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3549

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3548 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3489

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3488 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTGCT 3429

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3428 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3369

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3368 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3309

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3308 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3249

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3248 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3189

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3188 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3129

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3128 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3069

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3068 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3009

Qy   3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3008 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 2949

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   2948 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 2889

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   2888 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 2829

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   2828 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 2769

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   2768 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 2709

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   2708 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 2649

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   2648 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 2589

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   2588 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 2529

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   2528 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 2469

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db   2468 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 2409

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   2408 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 2349

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   2348 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 2289

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   2288 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 2229

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   2228 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 2169

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   2168 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 2109

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   2108 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 2049
```

```
Qy     4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db     2048 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 1989

Qy     4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db     1988 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 1929

Qy     4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db     1928 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 1869

Qy     4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db     1868 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGGAGGAT 1815

Qy     4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db     1814 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAGCAGTAAACTTGTCCAAAAC 1755

Qy     4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db     1754 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 1695

Qy     4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db     1694 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 1635

Qy     4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db     1634 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 1575

Qy     4721 GCAGCCCA---------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db     1574 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 1515

Qy     4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db     1514 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 1455

Qy     4819 ------------------------------------------------------------ 4818
Db     1454 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 1395

Qy     4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db     1394 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 1335

Qy     4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db     1334 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 1275

Qy     4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db     1274 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 1215

Qy     4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db     1214 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 1155

Qy     5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db     1154 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 1095

Qy     5079 GACCTGGCAGAATCCTAAGATGTTAGTTG---------------------CTCATTGTAC 5117
Db     1094 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 1035

Qy     5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db     1034 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 976

Qy     5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db      975 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 917

Qy     5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db      916 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 858

Qy     5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db      857 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 818

Qy     5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db      817 ----------------------------------------CATATAAATTCTGATTTA 800

Qy     5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db      799 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 740

Qy     5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db      739 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 683

Qy     5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db      682 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 623

Qy     5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db      622 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 563

Qy     5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db      562 TATAACCCACGAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 503

Qy     5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db      502 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 443

Qy     5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATATT 5806
Db      442 ATTACACGGAGTGTGCTAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 383

Qy     5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db      382 ATTGGGAAATTT-------TTCTTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 330

Qy     5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db      329 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 270

Qy     5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db      269 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 210

Qy     5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db      209 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 152

Qy     6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db      151 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 92

Qy     6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db       91 TGCCATTCTTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 32

Qy     6134 CTGATTTAAATAAA 6147
Db       31 CTGATTTAAATAAA 18
```

RESULT 15

```
US-60-324-185-30580
; Sequence 30580, Application US/60324185
; GENERAL INFORMATION:
;  APPLICANT: Morris, MacDonald
;  APPLICANT:  Lal, Preeti
;  APPLICANT:  Diep, Dinh
;  TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
;  TITLE OF INVENTION:  POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
;  TITLE OF INVENTION:  POLYMORPHISMS IDENTIFIED THEREBY
;  FILE REFERENCE: GX-0019-1 P
;  CURRENT APPLICATION NUMBER: US/60/324,185
;  CURRENT FILING DATE:  2001-09-21
;  NUMBER OF SEQ ID NOS: 35862
;  SOFTWARE: PERL Program
; SEQ ID NO 30580
;   LENGTH: 6632
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 445786.20
US-60-324-185-30580

  Query Match             35.3%;  Score 2172.8;  DB 76;  Length 6632;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 4220;  Conservative    0;  Mismatches 1797;  Indels  539;  Gaps   43;

Qy        3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db      190 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 249

Qy       63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db      250 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 309

Qy      120 GGGCCC----------GCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGA 169
Db      310 GGCCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGA 369

Qy      170 CCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGA 229
Db      370 CCCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGA 429

Qy      230 ACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268
Db      430 GCAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGG 489

Qy      269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCG 307
Db      490 TCAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACAGAGGTTGGACAGAG 549

Qy      308 AGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGC 367
Db      550 AGACTCTGAAGATGTGAGCAAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGT 609

Qy      368 TGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCC 427
Db      610 TGTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCC 669

Qy      428 TGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGT 487
Db      670 TTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATAT 729

Qy      488 TGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAA 547
Db      730 TGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGAC 789

Qy      548 TGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGA 607
Db      790 AGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGC 849

Qy      608 AGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTC 667
Db      850 A---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGC 894

Qy      668 AGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCA 727
Db      895 AGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCG 954

Qy      728 AGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGA 787
Db      955 TGAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAA---GCAGTGGAGGA 1011

Qy      788 AGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCC 847
Db     1012 ATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCC 1071

Qy      848 GAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTG 907
Db     1072 GACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTG 1131

Qy      908 GGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGC 967
Db     1132 GGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTC 1191

Qy      968 CGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGA 1027
Db     1192 AGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGA 1251

Qy     1028 GCCAGCCTCGGAGGA-------------------------GCAGGAGCCGGCAGAAGA 1060
Db     1252 GGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAG 1311

Qy     1061 CACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCA 1120
Db     1312 TGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCA 1371

Qy     1121 GGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTT 1180
Db     1372 AGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTT 1431

Qy     1181 TGATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGT 1237
Db     1432 TGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGT 1491

Qy     1238 GGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGA 1297
Db     1492 GGAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGA 1527

Qy     1298 AGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGA 1357
Db     1528 AGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGA 1587

Qy     1358 AGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCA 1417
Db     1588 AGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCC 1647

Qy     1418 CACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCAT 1477
Db     1648 TACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGT 1707

Qy     1478 TGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTT 1537
Db     1708 TGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACT 1767

Qy     1538 GAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAA 1597
Db     1768 AAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAA 1827

Qy     1598 ACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCC 1657
Db     1828 AAGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCC 1884

Qy     1658 AGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGA 1717
Db     1885 GGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGA 1944
```

```
Qy   1718 GACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGG 1777
Db   1945 GATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGG 2004

Qy   1778 AACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAA 1837
Db   2005 AGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAA 2064

Qy   1838 GATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCT 1897
Db   2065 GATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCT 2124

Qy   1898 GGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGA 1957
Db   2125 GGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGA 2184

Qy   1958 TGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATAC 2017
Db   2185 AGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATAC 2244

Qy   2018 TTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGC 2077
Db   2245 CTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAG 2304

Qy   2078 ATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGA 2137
Db   2305 GTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGA 2364

Qy   2138 GGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCA 2197
Db   2365 TGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACA 2424

Qy   2198 GGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGG 2257
Db   2425 TGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGG 2484

Qy   2258 TGTCTCCACTTGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAA 2317
Db   2485 CGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAA 2544

Qy   2318 ACTGGAAGAGAAAGCCGAAGAC-------------TCTAGTGTAGAGCAGTTGTCCACTGA 2365
Db   2545 GCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGA 2604

Qy   2366 GATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCG 2425
Db   2605 CACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAG 2664

Qy   2426 GAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGA 2485
Db   2665 GAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGG 2724

Qy   2486 GATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGT 2545
Db   2725 GGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGT 2784

Qy   2546 GGAGAGGGAGAAGATGGAAGCCCAG----------GGGAATACGGAGCTGCCCCAGCTGC 2595
Db   2785 AGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAAGCGCAGAGCAGCCCGAGCAGA 2844

Qy   2596 TGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCG 2655
Db   2845 AGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAG 2904

Qy   2656 CAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATAT 2715
Db   2905 CTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATAT 2964

Qy   2716 CCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAG 2775
Db   2965 CTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGG 3024

Qy   2776 AGGTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTAC 2832
Db   3025 AGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGC 3084

Qy   2833 CAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAG 2892
Db   3085 CAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAG 3144

Qy   2893 CTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGG 2952
Db   3145 CTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTG 3204

Qy   2953 GGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACA 3012
Db   3205 CTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACA 3264

Qy   3013 CCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAG 3072
Db   3265 CCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGC 3324

Qy   3073 AGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCC 3132
Db   3325 AAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCC 3384

Qy   3133 AGGTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTG 3189
Db   3385 AGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAG 3444

Qy   3190 AGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGA 3249
Db   3445 AAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGA 3504

Qy   3250 AAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGG 3309
Db   3505 AAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGA 3564

Qy   3310 CTACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------C 3357
Db   3565 CCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGC 3624

Qy   3358 ATGTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGG 3399
Db   3625 TTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGG 3684

Qy   3400 AACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCA 3459
Db   3685 AACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCA 3744

Qy   3460 CTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCC 3519
Db   3745 CCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCC 3804

Qy   3520 AGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTA 3579
Db   3805 ATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTC 3864

Qy   3580 CTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATG 3639
Db   3865 CTGCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAG 3924

Qy   3640 GGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCG 3699
Db   3925 AACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTG 3984

Qy   3700 TGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGT 3750
Db   3985 TATCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAA 4044

Qy   3751 TGGATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCA 3804
Db   4045 CCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCA 4104

Qy   3805 ACAGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTC 3864
```

```
Db   4105 GTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTA 4164

Qy   3865 AGGAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGG 3909
Db   4165 AAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGA 4224

Qy   3910 GAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTG 3969
Db   4225 GAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATG 4284

Qy   3970 AAGAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAG 4026
Db   4285 AAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCC 4344

Qy   4027 TCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAA 4086
Db   4345 TCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAA 4404

Qy   4087 GCCCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCC 4143
Db   4405 GCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCT 4464

Qy   4144 TGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGG 4197
Db   4465 CTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTG 4524

Qy   4198 TTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTG 4257
Db   4525 CCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAG 4584

Qy   4258 AGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCC 4317
Db   4585 AGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCA 4644

Qy   4318 TGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCA 4377
Db   4645 CAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAG 4704

Qy   4378 CCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAG 4437
Db   4705 GCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAG 4764

Qy   4438 AGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAG 4497
Db   4765 TCGATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGGAGG 4818

Qy   4498 TCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGA 4557
Db   4819 ATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAA 4878

Qy   4558 ACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTC 4614
Db   4879 ACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGT 4938

Qy   4615 ATGCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGAC 4658
Db   4939 TGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGAC 4998

Qy   4659 AGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTG 4718
Db   4999 AGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTA 5058

Qy   4719 CCGCAGCCCA---------------------GAGAGGACTTGCAAGTCCTGACCGTTCTG 4757
Db   5059 CTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCA 5118

Qy   4758 GAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCA 4817
Db   5119 AAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAA 5178

Qy   4818 A----------------------------------------------------------- 4818

Db   5179 ATGATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAA 5238

Qy   4819 ----------------------------------------AGTAAGCATTGAGAAGCTGC 4838
Db   5239 CAAAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCAC 5298

Qy   4839 CTCCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCT 4896
Db   5299 TAGTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACT 5358

Qy   4897 TAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACG 4956
Db   5359 CAGCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATG 5418

Qy   4957 GACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGT 5016
Db   5419 GACCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGC 5478

Qy   5017 CTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGG 5076
Db   5479 ACAGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGA 5538

Qy   5077 GAGACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGT 5115
Db   5539 GAGAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGT 5598

Qy   5116 ACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGG 5171
Db   5599 CTGTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAG 5658

Qy   5172 ACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTT 5231
Db   5659 A-CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCT 5716

Qy   5232 CCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACA 5291
Db   5717 CATTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACA 5775

Qy   5292 TTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACA 5351
Db   5776 TTTCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC------------------ 5817

Qy   5352 TTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATT 5411
Db   5818 --------------------------------------------CATATAAATTCTGATT 5833

Qy   5412 TAAGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAA 5470
Db   5834 TAAGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAA 5893

Qy   5471 CTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCT 5530
Db   5894 CTGGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCT 5950

Qy   5531 ATTGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTAT 5573
Db   5951 ATTGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTAT 6010

Qy   5574 TATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAAC 5633
Db   6011 CACACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGT 6070

Qy   5634 TGTAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTT 5684
Db   6071 TATATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTC 6130

Qy   5685 TTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTC 5744
Db   6131 TTTAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTC 6190

Qy   5745 CTATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATA 5804
Db   6191 CTATTACACGGAGTGTGCTAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTG 6250
```

```
Qy   5805 TTACCATGATTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTC 5864

Db   6251 TCATTGGGAAATTT-------TTCTTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATT 6303

Qy   5865 TGTGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGA 5907

Db   6304 CTGGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAA 6363

Qy   5908 GTGTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCT 5951

Db   6364 GTGTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATT 6423

Qy   5952 GTTGTACGTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAG 6011

Db   6424 GTTGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAA 6481

Qy   6012 TGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATC 6071

Db   6482 TAAGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATT 6541

Qy   6072 TGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGAC 6131

Db   6542 TGTGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGAC 6601

Qy   6132 AACTGATTTAAATAAA 6147

Db   6602 GACTGATTTAAATAAA 6617
```

Search completed: December 13, 2002, 00:00:20
Job time : 11603 secs

```
OM nucleic - nucleic search, using sw model

Run on:         December 12, 2002, 14:19:40 ; Search time 391 Seconds
                                          (without alignments)
                                          11377.814 Million cell updates/sec

Title:          US-09-902-432-3
Perfect score:  6160
Sequence:       1 cgggccggcgggagtagaag..........aaataaacatatttgactac 6160

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       416730 seqs, 361097816 residues

Total number of hits satisfying chosen parameters:      833460

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Pending_Patents_NA_New:*
                1:  /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
                2:  /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
                3:  /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
                4:  /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
                5:  /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
                6:  /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
                7:  /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	5204.2	84.5	5236	6	US-10-152-319A-2063	Sequence 2063, Ap
	2	2144.4	34.8	6608	1	PCT-US02-04915-43	Sequence 43, Appl
	3	1994.8	32.4	5931	5	US-09-724-676-12735	Sequence 12735, A
	4	1994.8	32.4	5931	5	US-09-724-676A-12735	Sequence 12735, A
	5	1920.8	31.2	5773	5	US-09-724-676-12736	Sequence 12736, A
	6	1920.8	31.2	5773	5	US-09-724-676A-12736	Sequence 12736, A
c	7	550	8.9	563	6	US-10-152-319A-1312	Sequence 1312, Ap
	8	162.2	2.6	178	6	US-10-152-319A-2137	Sequence 2137, Ap
c	9	124.6	2.0	180	6	US-10-152-319A-1193	Sequence 1193, Ap
	10	119	1.9	753	5	US-09-724-676-12734	Sequence 12734, A
	11	119	1.9	753	5	US-09-724-676A-12734	Sequence 12734, A
	12	71	1.2	575	6	US-10-203-138A-9167	Sequence 9167, Ap
	13	71	1.2	1969	6	US-10-203-138A-4036	Sequence 4036, Ap
	14	70.8	1.1	1926	6	US-10-194-046-3	Sequence 3, Appli
c	15	70.8	1.1	8705	6	US-10-273-678-16	Sequence 16, Appl
c	16	70.8	1.1	8705	6	US-10-291-230-14	Sequence 14, Appl
c	17	70.8	1.1	8705	6	US-10-291-249-14	Sequence 14, Appl
c	18	70.8	1.1	16080	5	US-09-724-566A-48	Sequence 48, Appl
	19	68.4	1.1	107	5	US-09-620-607B-815	Sequence 815, App
	20	66.4	1.1	3489	6	US-10-194-046-1	Sequence 1, Appli
c	21	65.8	1.1	659158	5	US-09-771-208A-20	Sequence 20, Appl
	22	61	1.0	700	6	US-10-203-138A-5915	Sequence 5915, Ap
	23	56.4	0.9	2550	5	US-09-724-676-41735	Sequence 41735, A
	24	56.4	0.9	2550	5	US-09-724-676A-41735	Sequence 41735, A
	25	56.4	0.9	2709	5	US-09-724-676-41736	Sequence 41736, A
	26	56.4	0.9	2709	5	US-09-724-676A-41736	Sequence 41736, A
	27	56.4	0.9	2805	5	US-09-724-676-41728	Sequence 41728, A
	28	56.4	0.9	2805	5	US-09-724-676A-41728	Sequence 41728, A
	29	56.4	0.9	2964	5	US-09-724-676-41729	Sequence 41729, A
	30	56.4	0.9	2964	5	US-09-724-676A-41729	Sequence 41729, A
	31	56.4	0.9	3382	5	US-09-724-676-41723	Sequence 41723, A
	32	56.4	0.9	3382	5	US-09-724-676A-41723	Sequence 41723, A
	33	56.4	0.9	3637	5	US-09-724-676-41717	Sequence 41717, A
	34	56.4	0.9	3637	5	US-09-724-676A-41717	Sequence 41717, A
	35	56.4	0.9	3817	5	US-09-724-676-41745	Sequence 41745, A
	36	56.4	0.9	3817	5	US-09-724-676A-41745	Sequence 41745, A
	37	56.4	0.9	4072	5	US-09-724-676-41732	Sequence 41732, A
	38	56.4	0.9	4072	5	US-09-724-676A-41732	Sequence 41732, A
	39	56.2	0.9	474	6	US-10-203-138A-3499	Sequence 3499, Ap
	40	56.2	0.9	906	6	US-10-273-334-33	Sequence 33, Appl
	41	55.4	0.9	2142	6	US-10-152-319A-1533	Sequence 1533, Ap
	42	55	0.9	905	6	US-10-273-334-11	Sequence 11, Appl
	43	55	0.9	907	6	US-10-273-334-9	Sequence 9, Appli
	44	55	0.9	907	6	US-10-273-334-13	Sequence 13, Appl
	45	55	0.9	907	6	US-10-273-334-25	Sequence 25, Appl

ALIGNMENTS

```
RESULT 1
US-10-152-319A-2063
; Sequence 2063, Application US/10152319A
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Higgs, Brandon
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5089-US
;  CURRENT APPLICATION NUMBER: US/10/152,319A
;  CURRENT FILING DATE:  2002-05-22
;  PRIOR APPLICATION NUMBER: US 60/292,335
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/297,523
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,925
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,810
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,807
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,808
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/315,047
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/324,928
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/330,867
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/330,462
;  PRIOR FILING DATE: 2001-10-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 2221
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2063
;   LENGTH: 5236
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. NM_057103
US-10-152-319A-2063

  Query Match           84.5%;  Score 5204.2;  DB 6;  Length 5236;
  Best Local Similarity  99.7%;  Pred. No. 0;
  Matches 5223;  Conservative    0;  Mismatches   13;  Indels    1;  Gaps    1;
```

```
Qy          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy        361 CCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600

Qy        601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660

Qy        661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy        721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy        781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy        841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy        901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy        961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy       1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy       1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy       1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy       1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy       1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy       1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy       1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy       1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy       1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy       1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy       1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy       1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy       1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy       1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy       1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy       1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy       1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy       2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy       2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy       2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy   2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy   2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy   2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400

Qy   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
          |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAGAACAAGCCA 2460

Qy   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
          |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTGATGTCCCAGCCGTCG 2520

Qy   2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
          ||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |||
Db   2521 TGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATGCGG 2580

Qy   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760

Qy   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120

Qy   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180

Qy   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
```

```
Qy    4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy    4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy    4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy    4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy    4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
           ||||||||||||||||||||||||      |  ||||||||||||||||||||||||||
Db    4621 ATGATTCACAGACCCAGGTTCCTGCAATGCAGGCTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy    4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
           |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    4681 GGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy    4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy    4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGAT-CAAA 4859

Qy    4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4860 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4919

Qy    4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4920 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4979

Qy    4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4980 ATCGCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5039

Qy    5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5040 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5099

Qy    5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5100 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5159

Qy    5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5160 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5219

Qy    5221 TCCAATGATTTCCACCC 5237
           |||||||||||||||||
Db    5220 TCCAATGATTTCCACCC 5236
```

RESULT 2
PCT-US02-04915-43
; Sequence 43, Application PC/TUS0204915
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Glynne, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Aziz, Natasha
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
; TITLE OF INVENTION: Methods of Screening for Angiogenesis Modulators
; FILE REFERENCE: 018501-006200PC
; CURRENT APPLICATION NUMBER: PCT/US02/04915
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: US 09/784,356
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: US 09/791,390
; PRIOR FILING DATE: 2001-02-22
; PRIOR APPLICATION NUMBER: US 60/285,475
; PRIOR FILING DATE: 2002-04-19
; PRIOR APPLICATION NUMBER: US 60/310,025
; PRIOR FILING DATE: 2001-08-03
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/334,244
; PRIOR FILING DATE: 2001-11-29
; NUMBER OF SEQ ID NOS: 230
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
;   LENGTH: 6608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
PCT-US02-04915-43

```
  Query Match             34.8%;  Score 2144.4;  DB 1;  Length 6608;
  Best Local Similarity   64.3%;  Pred. No. 0;
  Matches 4215;  Conservative    0;  Mismatches 1796;  Indels  543;  Gaps   45;

Qy       3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
           |||  |   |   | |||||  ||       |||||||||||| || || |||||||||||||
Db     162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy      63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
           ||||| |||||||||    | |||||||  ||||||  ||||   | || ||| |||
Db     222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy     120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
           || ||| | || || || ||  |              || |      | |  ||| | | |||||
Db     282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy     171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
           |||||||||||||||| ||||||||||| ||||||||  |  |||| |||||||||||
Db     342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy     231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268
           ||||  ||  ||    |||| ||||    |||    | ||
Db     402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy     269 -------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
                               | |||| ||||||||| |||    ||| ||||||||| ||
Db     462 CAAGGAGCCCTAAACAGCCAGCAGGAGGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy     309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
           || || ||||||||||| |||| ||||   | |||| |||||  | || ||  |  |
Db     522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy     369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
           ||| | || |||||| | ||||||||||||||          ||   || |||||||| |||
Db     582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy     429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
            |||||||| |||| | |||||   |    ||| ||  |||||||||||||||||||| ||
Db     639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy     489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
           || || ||||| || ||||| |||||||| |||||||||| ||||| ||||||||  |
Db     699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy     549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
           || |||  | || ||||||| ||||||||||| ||||| |||||||| || || |||| |
Db     759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy     609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
                    || ||||| |||||||||| |||||  | | | ||      ||||| ||
Db     819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863
```

```
Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1461 GAGGAGAGAACCGAAGAGCAGAAA------------------------ACGGAGGTGGAA 1496

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
```

```
Db  2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy  2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db  3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy  2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db  3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy  2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db  3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy  3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db  3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy  3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db  3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy  3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db  3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy  3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db  3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy  3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db  3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy  3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db  3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy  3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db  3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy  3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db  3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy  3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db  3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy  3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db  3774 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3833

Qy  3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db  3834 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3893

Qy  3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db  3894 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3953

Qy  3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db  3954 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4013

Qy  3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db  4014 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4073

Qy  3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db  4074 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4133

Qy  3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db  4134 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4193

Qy  3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db  4194 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4253

Qy  3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db  4254 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4313

Qy  4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db  4314 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4373

Qy  4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db  4374 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4433

Qy  4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db  4434 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4493

Qy  4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db  4494 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4553

Qy  4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db  4554 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4613

Qy  4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db  4614 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4673

Qy  4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db  4674 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4733

Qy  4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db  4734 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4784

Qy  4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db  4785 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4844

Qy  4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db  4845 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4904

Qy  4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db  4905 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4964

Qy  4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db  4965 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5024

Qy  4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db  5025 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5084

Qy  4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db  5085 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5144

Qy  4819 ------------------------------------------------------------ 4818
Db  5145 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5204
```

```
Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5205 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5264

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5265 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5324

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5325 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5384

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5385 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5444

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5445 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5504

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5505 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5564

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5565 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5623

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5624 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5682

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5683 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5741

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5742 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5781

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5782 -------------------------------------------CATATAAATTCTGATTTA 5799

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5800 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5859

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5860 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5916

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5917 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5976

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5977 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6036

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6037 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6096

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6097 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6156

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6157 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6216

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6217 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6269

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6270 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6329

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6330 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6389

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6390 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6447

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6448 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6507

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6508 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6567

Qy   6134 CTGATTTAAATAAA 6147
Db   6568 CTGATTTAAATAAA 6581


RESULT 3
US-09-724-676-12735
; Sequence 12735, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12735
;   LENGTH: 5931
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-12735

  Query Match             32.4%;  Score 1994.8;  DB 5;  Length 5931;
  Best Local Similarity   67.6%;  Pred. No. 0;
  Matches 3281;  Conservative    0;  Mismatches 1332;  Indels  237;  Gaps   25;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy    269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACAGAGGTTGGACAGAGA 521
```

```
Qy   309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db   522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy   369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db   582 GTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCT 641

Qy   429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db   642 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 701

Qy   489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db   702 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 761

Qy   549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db   762 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 821

Qy   609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db   822 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 866

Qy   669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db   867 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 926

Qy   729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db   927 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAA---GCAGTGGAGGAA 983

Qy   789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db   984 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1043

Qy   849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db  1044 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1103

Qy   909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db  1104 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1163

Qy   969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db  1164 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1223

Qy  1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db  1224 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1283

Qy  1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db  1284 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1343

Qy  1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db  1344 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1403

Qy  1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db  1404 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1463

Qy  1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db  1464 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1499

Qy  1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db  1500 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAA 1559

Qy  1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db  1560 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1619

Qy  1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db  1620 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1679

Qy  1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db  1680 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1739

Qy  1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db  1740 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGAAAA 1799

Qy  1599 CGAGGAGGTGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db  1800 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1856

Qy  1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db  1857 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1916

Qy  1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db  1917 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1976

Qy  1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db  1977 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2036

Qy  1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db  2037 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2096

Qy  1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db  2097 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2156

Qy  1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db  2157 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2216

Qy  2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db  2217 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGG 2276

Qy  2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db  2277 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2336

Qy  2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db  2337 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2396

Qy  2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db  2397 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2456

Qy  2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db  2457 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2516

Qy  2319 CTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db  2517 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2576

Qy  2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db  2577 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2636

Qy  2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
```

```
Db   2637 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2696

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2697 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2756

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2757 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAG 2816

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2817 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2876

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2877 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2936

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2937 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2996

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2997 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3056

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3057 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3116

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3117 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTGCT 3176

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3177 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3236

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3237 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3296

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3297 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3356

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3357 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3416

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3417 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3476

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3477 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3536

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTA---------GACCAT 3359
Db   3537 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3596

Qy   3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3597 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3656

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3657 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3716

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3717 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3776

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3777 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3836

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3837 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3896

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3897 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3956

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3957 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4016

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4017 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4076

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4077 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4136

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4137 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4196

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4197 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4256

Qy   3972 G---AAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4257 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4316

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4317 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4376

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4377 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4436

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCAT------AGAAACGGTT 4199
Db   4437 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4496

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4497 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4556

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4557 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4616

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4617 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4676

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4677 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4736

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4737 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGGAGGAT 4790

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4791 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4850
```

```
Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4851 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4910

Qy   4617 GCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGA 4666
Db   4911 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGA 4960


RESULT 4
US-09-724-676A-12735
; Sequence 12735, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12735
;   LENGTH: 5931
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676A-12735

  Query Match             32.4%;  Score 1994.8;  DB 5;  Length 5931;
  Best Local Similarity   67.6%;  Pred. No. 0;
  Matches 3281;  Conservative    0;  Mismatches 1332;  Indels  237;  Gaps   25;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy    269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACAGAGGTTGGACAGAGA 521

Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    582 GTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCT 641

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    642 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 701

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    702 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 761

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    762 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 821

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    822 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 866

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    867 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 926

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    927 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAA---GCAGTGGAGGAA 983

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    984 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1043

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1044 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1103

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1104 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1163

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1164 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1223

Qy   1029 CCAGCCTCGGAGGA-------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1224 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1283

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1284 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1343

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1344 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1403

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1404 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1463

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1464 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1499

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1500 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAA 1559

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1560 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1619

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1620 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1679

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1680 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1739

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1740 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1799

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1800 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1856
```

```
Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1857 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1916

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1917 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1976

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1977 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2036

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2037 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2096

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2097 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2156

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2157 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2216

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2217 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGG 2276

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGAGGGGGACAGTCACAGAGCAGAG 2138
Db   2277 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2336

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2337 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2396

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2397 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2456

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2457 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2516

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2517 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2576

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2577 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2636

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2637 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2696

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2697 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2756

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2757 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAG 2816

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2817 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2876

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2877 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2936

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2937 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2996

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2997 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3056

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3057 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3116

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3117 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTGCT 3176

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3177 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3236

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3237 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3296

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3297 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3356

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3357 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3416

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3417 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3476

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3477 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3536

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTA---------GACCAT 3359
Db   3537 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3596

Qy   3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3597 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3656

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3657 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3716

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3717 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3776

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3777 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3836

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3837 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3896

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3897 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3956

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3957 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4016

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
```

```
Db   4017 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4076

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4077 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4136

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db   4137 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4196

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4197 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4256

Qy   3972 G---AAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4257 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4316

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4317 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4376

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4377 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4436

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCAT------AGAAACGGTT 4199
Db   4437 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4496

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4497 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4556

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4557 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4616

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4617 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4676

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4677 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4736

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4737 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGGAGGAT 4790

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4791 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4850

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4851 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4910

Qy   4617 GCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGA 4666
Db   4911 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGA 4960
```

RESULT 5
US-09-724-676-12736
; Sequence 12736, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12736
; LENGTH: 5773
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (91)..(91)
; OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-12736

```
  Query Match             31.2%;  Score 1920.8;  DB 5;  Length 5773;
  Best Local Similarity   68.2%;  Pred. No. 0;
  Matches 3072;  Conservative    0;  Mismatches 1247;  Indels  183;  Gaps   22;

Qy    297 GTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAGACCGAGTTGAAGAAATGGCGGCC 356
Db    352 GTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACT 411

Qy    357 AACTCCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATT 416
Db    412 AAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATC 471

Qy    417 GAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAG 476
Db    472 GAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAG 531

Qy    477 GCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAG 536
Db    532 GCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAA 591

Qy    537 AAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGC 596
Db    592 AAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGG 651

Qy    597 GAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCC 656
Db    652 GAGGGAGCAGCA---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT--- 699

Qy    657 GTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGC 716
Db    700 ---GGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAG 756

Qy    717 ACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCG 776
Db    757 ACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCAGCCGAATCTGGCCAA--- 813

Qy    777 GCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCC 836
Db    814 GCAGTGGAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCT 873

Qy    837 CCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTC 896
Db    874 GCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTC 933

Qy    897 ACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGAT 956
Db    934 ACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAA 993

Qy    957 CTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAG 1016
Db    994 GTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGAC 1053

Qy   1017 GAAAAGACAGAGCCAGCCTCGGAGGA--------------------------GCAGGAG 1049
Db   1054 GGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAG 1113

Qy   1050 CCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCT 1109
Db   1114 CCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCC 1173

Qy   1110 TTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCA 1169
```

```
Db   1174 TCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCG 1233

Qy   1170 ACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCAC 1226
Db   1234 ACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCAC 1293

Qy   1227 GTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGC 1286
Db   1294 GTCAGCACCGTGGAGGAGAGAACCGAAGAGCAGAAA------------------------ 1329

Qy   1287 GTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAG 1346
Db   1330 ACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCA 1389

Qy   1347 GTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCT 1406
Db   1390 GAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCC 1449

Qy   1407 GGAGGAGACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACAC 1466
Db   1450 GGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCC 1509

Qy   1467 CCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAG 1526
Db   1510 CCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAG 1569

Qy   1527 GGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAG 1586
Db   1570 GGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAA 1629

Qy   1587 CAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCAC 1646
Db   1630 CAGAAAGGGAAAAGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCA 1686

Qy   1647 ACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAG 1706
Db   1687 GCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAG 1746

Qy   1707 GAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAA 1766
Db   1747 GAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAA 1806

Qy   1767 GCTGAGGAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCA 1826
Db   1807 GCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCA 1866

Qy   1827 TCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAG 1886
Db   1867 TCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAA 1926

Qy   1887 GAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCA 1946
Db   1927 GAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCT 1986

Qy   1947 GAAATGCAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGT 2006
Db   1987 GAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGC 2046

Qy   2007 AGGGTGGATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGA 2066
Db   2047 AAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGA 2106

Qy   2067 GCAAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGT 2126
Db   2107 GCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCAC 2166

Qy   2127 CACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGC 2186
Db   2167 CAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGT 2226

Qy   2187 ACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCC 2246
Db   2227 TCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACC 2286

Qy   2247 GAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAA 2306
Db   2287 GAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAA 2346

Qy   2307 TCCAAGTCAAAACTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAG 2354
Db   2347 TCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACAT 2406

Qy   2355 TTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATC 2414
Db   2407 TCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATT 2466

Qy   2415 CCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCA 2474
Db   2467 CCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCA 2526

Qy   2475 GGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAG 2534
Db   2527 GGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAG 2586

Qy   2535 TATAATGCAGTGGAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTG 2585
Db   2587 TATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAG 2646

Qy   2586 CCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACT 2645
Db   2647 CCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATG 2706

Qy   2646 GTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCT 2705
Db   2707 ATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCT 2766

Qy   2706 TCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCA 2765
Db   2767 TCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTG 2826

Qy   2766 CCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCA---GAAGAAACTCCTGTGCTCACC 2822
Db   2827 TTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACT 2886

Qy   2823 CAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTC 2882
Db   2887 GAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTG 2946

Qy   2883 ACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACC 2942
Db   2947 ACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACC 3006

Qy   2943 GAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGAC 3002
Db   3007 GAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGAC 3066

Qy   3003 TCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGAT 3062
Db   3067 TCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGAC 3126

Qy   3063 ACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAA 3122
Db   3127 ATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAA 3186

Qy   3123 GAGGAGTCCCAGGTGCCTGCAACCCA---GACTGTGCAGAGAACGGGGTCAAAAGCACTG 3179
Db   3187 GAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGA 3246

Qy   3180 GAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGAC 3239
Db   3247 GCAGAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGAT 3306

Qy   3240 GTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAG 3299
Db   3307 GTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTG 3366
```

```
Qy   3300 ACTGGACAGGCTACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTA--- 3353
Db   3367 GTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAG 3426

Qy   3354 ------GACCATGTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAG 3389
Db   3427 TCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAG 3486

Qy   3390 CAGCTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACA 3449
Db   3487 ATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACT 3546

Qy   3450 AATGGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACC 3509
Db   3547 GATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATT 3606

Qy   3510 ATTGACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAA 3569
Db   3607 GTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCA 3666

Qy   3570 GAGGCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAG 3629
Db   3667 GAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAA 3726

Qy   3630 GAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACT 3689
Db   3727 GAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCA 3786

Qy   3690 GCTGCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAG--------- 3740
Db   3787 GTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCT 3846

Qy   3741 GTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG- 3799
Db   3847 GATGAGAAAACCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGC 3906

Qy   3800 -----ACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTG 3854
Db   3907 ATAACAGTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAA 3966

Qy   3855 GCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGT---------------CTTAGC 3899
Db   3967 GCTGAATGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCC 4026

Qy   3900 CTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAG 3959
Db   4027 CCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACC 4086

Qy   3960 CAAGTGAGTGAAG---AAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTAC 4016
Db   4087 CATGTGAATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGT 4146

Qy   4017 GGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGC 4076
Db   4147 AAGCAGCTCCTCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGT 4206

Qy   4077 CTTGGAGGAAGCCCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAA 4133
Db   4207 TTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAA 4266

Qy   4134 GTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCAT---- 4189
Db   4267 GTTCAGAGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTA 4326

Qy   4190 --AGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTA 4247
Db   4327 GGAGAAACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTA 4386

Qy   4248 TTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGAC 4307
Db   4387 GTTCTGGAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGAT 4446

Qy   4308 ACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCT 4367
Db   4447 GCTGTGCCCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACT 4506

Qy   4368 CCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGA 4427
Db   4507 ACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAG 4566

Qy   4428 TCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCA 4487
Db   4567 TCAGATGAAGTCGATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCA 4620

Qy   4488 ATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAG 4547
Db   4621 ATTGAGGAGGATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAA 4680

Qy   4548 ATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCC 4604
Db   4681 CTTGTCCAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCC 4740

Qy   4605 CCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGG 4664
Db   4741 ACCGAAATGTTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAG 4800

Qy   4665 GA 4666
Db   4801 GA 4802
```

RESULT 6
US-09-724-676A-12736
; Sequence 12736, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12736
; LENGTH: 5773
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (91)..(91)
; OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-12736

```
  Query Match             31.2%;  Score 1920.8;  DB 5;  Length 5773;
  Best Local Similarity   68.2%;  Pred. No. 0;
  Matches 3072;  Conservative    0;  Mismatches 1247;  Indels  183;  Gaps   22;

Qy    297 GTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCC 356
Db    352 GTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACT 411

Qy    357 AACTCCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATT 416
Db    412 AAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATC 471

Qy    417 GAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAG 476
Db    472 GAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAG 531

Qy    477 GCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAG 536
Db    532 GCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAA 591

Qy    537 AAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGC 596
Db    592 AAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGG 651
```

```
Qy    597 GAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCC 656
Db    652 GAGGGAGCAGCA---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT--- 699

Qy    657 GTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGC 716
Db    700 ---GGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAG 756

Qy    717 ACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCG 776
Db    757 ACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAA--- 813

Qy    777 GCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCC 836
Db    814 GCAGTGGAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCT 873

Qy    837 CCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTC 896
Db    874 GCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTC 933

Qy    897 ACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGAT 956
Db    934 ACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAA 993

Qy    957 CTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAG 1016
Db    994 GTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGAC 1053

Qy   1017 GAAAAGACAGAGCCAGCCTCGGAGGA--------------------------GCAGGAG 1049
Db   1054 GGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAG 1113

Qy   1050 CCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCT 1109
Db   1114 CCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCC 1173

Qy   1110 TTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCA 1169
Db   1174 TCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCG 1233

Qy   1170 ACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCAC 1226
Db   1234 ACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCAC 1293

Qy   1227 GTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGC 1286
Db   1294 GTCAGCACCGTGGAGGAGAGAACCGAAGAGCAGAAA------------------------ 1329

Qy   1287 GTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAG 1346
Db   1330 ACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCA 1389

Qy   1347 GTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCT 1406
Db   1390 GAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCC 1449

Qy   1407 GGAGGAGACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACAC 1466
Db   1450 GGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCC 1509

Qy   1467 CCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAG 1526
Db   1510 CCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAG 1569

Qy   1527 GGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAG 1586
Db   1570 GGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAA 1629

Qy   1587 CAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCAC 1646
Db   1630 CAGAAAGGGAAAAGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCA 1686

Qy   1647 ACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAG 1706
Db   1687 GCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAG 1746

Qy   1707 GAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAA 1766
Db   1747 GAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAA 1806

Qy   1767 GCTGAGGAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCA 1826
Db   1807 GCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCA 1866

Qy   1827 TCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAG 1886
Db   1867 TCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAA 1926

Qy   1887 GAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCA 1946
Db   1927 GAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCT 1986

Qy   1947 GAAATGCAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGT 2006
Db   1987 GAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGC 2046

Qy   2007 AGGGTGGATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGA 2066
Db   2047 AAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGA 2106

Qy   2067 GCAAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGT 2126
Db   2107 GCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCAC 2166

Qy   2127 CACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGC 2186
Db   2167 CAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGT 2226

Qy   2187 ACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCC 2246
Db   2227 TCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACC 2286

Qy   2247 GAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAA 2306
Db   2287 GAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAA 2346

Qy   2307 TCCAAGTCAAAACTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAG 2354
Db   2347 TCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACAT 2406

Qy   2355 TTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATC 2414
Db   2407 TCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATT 2466

Qy   2415 CCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCA 2474
Db   2467 CCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCA 2526

Qy   2475 GGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAG 2534
Db   2527 GGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAG 2586

Qy   2535 TATAATGCAGTGGAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTG 2585
Db   2587 TATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAG 2646

Qy   2586 CCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACT 2645
Db   2647 CCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATG 2706

Qy   2646 GTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCT 2705
Db   2707 ATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCT 2766

Qy   2706 TCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCA 2765
```

```
Db  2767 TCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTG 2826

Qy  2766 CCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCA---GAAGAAACTCCTGTGCTCACC 2822
Db  2827 TTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACT 2886

Qy  2823 CAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTC 2882
Db  2887 GAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTG 2946

Qy  2883 ACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACC 2942
Db  2947 ACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACC 3006

Qy  2943 GAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGAC 3002
Db  3007 GAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGAC 3066

Qy  3003 TCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGAT 3062
Db  3067 TCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGAC 3126

Qy  3063 ACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAA 3122
Db  3127 ATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAA 3186

Qy  3123 GAGGAGTCCCAGGTGCCTGCAACCCA---GACTGTGCAGAGAACGGGGTCAAAAGCACTG 3179
Db  3187 GAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGA 3246

Qy  3180 GAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGAC 3239
Db  3247 GCAGAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGAT 3306

Qy  3240 GTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAG 3299
Db  3307 GTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTG 3366

Qy  3300 ACTGGACAGGCTACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTA--- 3353
Db  3367 GTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAG 3426

Qy  3354 ------GACCATGTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAG 3389
Db  3427 TCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAG 3486

Qy  3390 CAGCTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACA 3449
Db  3487 ATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACT 3546

Qy  3450 AATGGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACC 3509
Db  3547 GATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATT 3606

Qy  3510 ATTGACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAA 3569
Db  3607 GTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCA 3666

Qy  3570 GAGGCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAG 3629
Db  3667 GAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAA 3726

Qy  3630 GAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACT 3689
Db  3727 GAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCA 3786

Qy  3690 GCTGCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAG--------- 3740
Db  3787 GTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCT 3846

Qy  3741 GTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG- 3799
Db  3847 GATGAGAAAACCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGC 3906

Qy  3800 -----ACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTG 3854
Db  3907 ATAACAGTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAA 3966

Qy  3855 GCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGT--------------CTTAGC 3899
Db  3967 GCTGAATGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCC 4026

Qy  3900 CTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAG 3959
Db  4027 CCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACC 4086

Qy  3960 CAAGTGAGTGAAG---AAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTAC 4016
Db  4087 CATGTGAATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGT 4146

Qy  4017 GGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGC 4076
Db  4147 AAGCAGCTCCTCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGT 4206

Qy  4077 CTTGGAGGAAGCCCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAA 4133
Db  4207 TTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAA 4266

Qy  4134 GTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCAT---- 4189
Db  4267 GTTCAGAGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTA 4326

Qy  4190 --AGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTA 4247
Db  4327 GGAGAAACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTA 4386

Qy  4248 TTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGAC 4307
Db  4387 GTTCTGGAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGAT 4446

Qy  4308 ACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCT 4367
Db  4447 GCTGTGCCCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACT 4506

Qy  4368 CCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGA 4427
Db  4507 ACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAG 4566

Qy  4428 TCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCA 4487
Db  4567 TCAGATGAAGTCGATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCA 4620

Qy  4488 ATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAG 4547
Db  4621 ATTGAGGAGGATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAA 4680

Qy  4548 ATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCC 4604
Db  4681 CTTGTCCAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCC 4740

Qy  4605 CCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGG 4664
Db  4741 ACCGAAATGTTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAG 4800

Qy  4665 GA 4666
Db  4801 GA 4802
```

RESULT 7
US-10-152-319A-1312/c
; Sequence 1312, Application US/10152319A
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark

```
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Higgs, Brandon
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5089-US
;  CURRENT APPLICATION NUMBER: US/10/152,319A
;  CURRENT FILING DATE:  2002-05-22
;  PRIOR APPLICATION NUMBER: US 60/292,335
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/297,523
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,925
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,810
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,807
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,808
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/315,047
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/324,928
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/330,867
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/330,462
;  PRIOR FILING DATE: 2001-10-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 2221
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1312
;   LENGTH: 563
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. AI233818
US-10-152-319A-1312

  Query Match             8.9%;  Score 550;  DB 6;  Length 563;
  Best Local Similarity   99.1%;  Pred. No. 3.1e-120;
  Matches  553;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5601 ATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAGATAACATT 5660
              | ||||||||||||||||||||||||| ||||||||||||||||||||||| ||||||||
Db        563 AAAGCGACGCTTTGAAAAGCTCCAAGCTTCAACTGTAACCTGCAGCAAACAAATAACATT 504

Qy       5661 CCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCT 5720
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CCTGGCAAGAAAAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCT 444

Qy       5721 AGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTT 5780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 AGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTT 384

Qy       5781 TGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAG 5840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 TGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAG 324

Qy       5841 TCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGA 5900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGA 264

Qy       5901 CAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGT 5960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 CAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGT 204

Qy       5961 TTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTG 6020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 TTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTG 144

Qy       6021 TTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCT 6080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 TTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCT 84

Qy       6081 TCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 6140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 TCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 24

Qy       6141 AAATAAACATATTTGACT 6158
              ||||||||||||||| ||
Db         23 AAATAAACATATTTGCCT 6


RESULT 8
US-10-152-319A-2137
; Sequence 2137, Application US/10152319A
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Higgs, Brandon
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5089-US
;  CURRENT APPLICATION NUMBER: US/10/152,319A
;  CURRENT FILING DATE:  2002-05-22
;  PRIOR APPLICATION NUMBER: US 60/292,335
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/297,523
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,925
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,810
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,807
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,808
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/315,047
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/324,928
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/330,867
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/330,462
;  PRIOR FILING DATE: 2001-10-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 2221
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2137
;   LENGTH: 178
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. U75404
US-10-152-319A-2137

  Query Match             2.6%;  Score 162.2;  DB 6;  Length 178;
  Best Local Similarity   97.8%;  Pred. No. 8.3e-29;
  Matches  175;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy       5962 TTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTGT 6021
              |||||||||||||||||||||||||||||||||||||||||||||||||| | ||||||
Db          1 TTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTCACTCCTGT 60

Qy       6022 TCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTT 6081
              |||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db         61 TCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGT-CCCCTT 119

Qy       6082 CTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 6140
              ||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db        120 CTTTAAGAACAATGTTGCATTACGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 178
```

```
RESULT 9
US-10-152-319A-1193/c
; Sequence 1193, Application US/10152319A
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Higgs, Brandon
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5089-US
;  CURRENT APPLICATION NUMBER: US/10/152,319A
;  CURRENT FILING DATE:  2002-05-22
;  PRIOR APPLICATION NUMBER: US 60/292,335
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/297,523
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,925
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,810
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,807
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,808
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/315,047
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/324,928
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/330,867
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/330,462
;  PRIOR FILING DATE: 2001-10-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 2221
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1193
;   LENGTH: 180
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. AI228494
US-10-152-319A-1193

  Query Match             2.0%;  Score 124.6;  DB 6;  Length 180;
  Best Local Similarity   83.0%;  Pred. No. 6.4e-20;
  Matches  142;  Conservative    0;  Mismatches   29;  Indels    0;  Gaps    0;

Qy   5987 GACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTGATAGA 6046
          | ||||||| | |||||||||||||||||||||||| |||||| | || ||||||| | |
Db    180 GGCCGAATTTAAATGCAGATCGAAGTGAGTCCTGTTTTTTACAAAAGGGATTTTGAAAAA 121

Qy   6047 TACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGT 6106
            ||||| |||||  || ||||  |  || |||||||| ||||| |||||||||||| ||| ||
Db    120 AACTGGGGTTTGGTTGGGTTAAAATTGGGCCCCTTTTTTAAAAACAATGTTGCCTTAAGT 61

Qy   6107 TCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGAC 6157
          ||| ||||| |||||| ||||| |||||||||| |||||||| ||||| |
Db     60 TCCCTTGGAAAAATTGGGATTTGCCAACTGATTTAATAAACAAATTTGCC 10

RESULT 10
US-09-724-676-12734
; Sequence 12734, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12734
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-12734

  Query Match             1.9%;  Score 119;  DB 5;  Length 753;
  Best Local Similarity   64.7%;  Pred. No. 2.4e-18;
  Matches  214;  Conservative    0;  Mismatches  105;  Indels   12;  Gaps    2;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
          |||  |   |  | ||||| ||      |||||||||||| || || |||||||||||||
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
          ||||| |||||||||    | |||||||  |||||||   ||||   | || ||| |||
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
          || ||| | || || ||  || |            || |      | |   ||| | | |||||
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
          ||||||||||||||| |||||||||||| ||||||||   |   |||| ||||||||||
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGAT 290
          ||||  ||  ||      |||| ||||    |||     | || ||| | | ||      | |
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy    291 GAGGATGTTGGACAGCGAGAGTCAGAAGATG 321
            | |  |       |  |  |   ||   ||||| |
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAG 492

RESULT 11
US-09-724-676A-12734
; Sequence 12734, Application US/09724676A
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676A
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12734
;   LENGTH: 753
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676A-12734

  Query Match             1.9%;  Score 119;  DB 5;  Length 753;
  Best Local Similarity   64.7%;  Pred. No. 2.4e-18;
  Matches  214;  Conservative    0;  Mismatches  105;  Indels   12;  Gaps    2;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
          |||  |   |  | ||||| ||      |||||||||||| || || |||||||||||||
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
          ||||| |||||||||    | |||||||  |||||||   ||||   | || ||| |||
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
          || ||| | || || ||  || |            || |      | |   ||| | | |||||
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
          ||||||||||||||| |||||||||||| ||||||||   |   |||| ||||||||||
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGAT 290
```

```
Db     402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy     291 GAGGATGTTGGACAGCGAGAGTCAGAAGATG 321

Db     462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAG 492


RESULT 12
US-10-203-138A-9167
; Sequence 9167, Application US/10203138A
; GENERAL INFORMATION:
;  APPLICANT: Molecular Dynamics, Inc.
;  APPLICANT:  Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  ANALYSIS OF GENE EXPRESSION IN BT 474
;  FILE REFERENCE: PB 0004 WO 8
;  CURRENT APPLICATION NUMBER: US/10/203,138A
;  CURRENT FILING DATE:  2002-08-02
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 04 February 2000 (04.02.00)
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 26 May 2000 (26.05.00)
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 03 August 2000 (03.08.00)
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 03 October 2000 (03.10.00)
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 27 September 2000 (27.09.00)
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 21 September 2000 (21.09.00)
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 30 June 2000 (30.06.00)
;  NUMBER OF SEQ ID NOS: 15438
;  SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 9167
;   LENGTH: 575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005822.1
;   FEATURE:
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
;   FEATURE:
;   OTHER INFORMATION: NT HIT: AF185255.1, EVALUE 4.80e+00
;   FEATURE:
;   OTHER INFORMATION: EST_HUMAN HIT: BE071749.1, EVALUE 5.80e+00
US-10-203-138A-9167

  Query Match             1.2%;  Score 71;  DB 6;  Length 575;
  Best Local Similarity   53.3%;  Pred. No. 4.8e-07;
  Matches  172;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy     969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAA-GAAAAGGAAAAGACAGA 1027

Db     212 GAGGAGGAAGAGGAGGAGGAAGAGGAGGAAGAAGAGGAGAAAGGAGGAGAAGAAGGAAGA 271

Qy    1028 GCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGA 1087

Db     272 GGATCAGGAAGAGGAGGAGGAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAGGA 331

Qy    1088 CTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGA 1147

Db     332 AGAGGAGAAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGA 391

Qy    1148 GGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGA 1207

Db     392 GGAGGAGGAAGAGAAGAAGGAAAAGGAGGAAGAGGAGGAGGAGGAAGAGGAGCAGAAGAA 451

Qy    1208 AGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGG 1267

Db     452 GGAAGAGGAGGAGGAAGAGGAAGAAGAGGAGAAGGAGAAAGAAGAGGAGGAGGAAGAGGA 511

Qy    1268 AGGAGAGGCTGAAGGGGGCGTGG 1290

Db     512 AGAGGAGGAGGAAGCTGGACTTG 534


RESULT 13
US-10-203-138A-4036
; Sequence 4036, Application US/10203138A
; GENERAL INFORMATION:
;  APPLICANT: Molecular Dynamics, Inc.
;  APPLICANT:  Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  ANALYSIS OF GENE EXPRESSION IN BT 474
;  FILE REFERENCE: PB 0004 WO 8
;  CURRENT APPLICATION NUMBER: US/10/203,138A
;  CURRENT FILING DATE:  2002-08-02
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 04 February 2000 (04.02.00)
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 26 May 2000 (26.05.00)
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 03 August 2000 (03.08.00)
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 03 October 2000 (03.10.00)
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 27 September 2000 (27.09.00)
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 21 September 2000 (21.09.00)
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 30 June 2000 (30.06.00)
;  NUMBER OF SEQ ID NOS: 15438
;  SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 4036
;   LENGTH: 1969
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005822.1
;   FEATURE:
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
US-10-203-138A-4036

  Query Match             1.2%;  Score 71;  DB 6;  Length 1969;
  Best Local Similarity   53.3%;  Pred. No. 7.9e-07;
  Matches  172;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy     969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAA-GAAAAGGAAAAGACAGA 1027

Db     554 GAGGAGGAAGAGGAGGAGGAAGAGGAGGAAGAAGAGGAGAAAGGAGGAGAAGAAGGAAGA 613

Qy    1028 GCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGA 1087

Db     614 GGATCAGGAAGAGGAGGAGGAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAGGA 673

Qy    1088 CTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGA 1147

Db     674 AGAGGAGAAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGA 733

Qy    1148 GGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGA 1207

Db     734 GGAGGAGGAAGAGAAGAAGGAAAAGGAGGAAGAGGAGGAGGAGGAAGAGGAGCAGAAGAA 793

Qy    1208 AGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGG 1267

Db     794 GGAAGAGGAGGAGGAAGAGGAAGAAGAGGAGAAGGAGAAAGAAGAGGAGGAGGAAGAGGA 853

Qy    1268 AGGAGAGGCTGAAGGGGGCGTGG 1290
```

```
Db    854 AGAGGAGGAGGAAGCTGGACTTG 876

RESULT 14
US-10-194-046-3
; Sequence 3, Application US/10194046
; GENERAL INFORMATION:
;  APPLICANT: Robertson, Erle S.
;  TITLE OF INVENTION: LANA Viral Protein Binding Sites
;  FILE REFERENCE: UM-07142
;  CURRENT APPLICATION NUMBER: US/10/194,046
;  CURRENT FILING DATE:  2002-07-11
;  PRIOR APPLICATION NUMBER: 09/410,399
;  PRIOR FILING DATE: 1999-10-01
;  NUMBER OF SEQ ID NOS: 22
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein-Barr virus
US-10-194-046-3

  Query Match             1.1%;  Score 70.8;  DB 6;  Length 1926;
  Best Local Similarity   45.3%;  Pred. No. 8.8e-07;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy    946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db    333 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 392

Qy   1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db    393 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 452

Qy   1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db    453 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 512

Qy   1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db    513 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 572

Qy   1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db    573 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 632

Qy   1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db    633 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 691

Qy   1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db    692 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 749

Qy   1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db    750 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 809

Qy   1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db    810 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 869

Qy   1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db    870 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 929

Qy   1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
Db    930 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 989

Qy   1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665
Db    990 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 1049

Qy   1666 CTGATGAGCAGAAGGGAGAG 1685
Db   1050 TGGAGGCCGCCGGGGTAGAG 1069

RESULT 15
US-10-273-678-16/c
; Sequence 16, Application US/10273678
; GENERAL INFORMATION:
;  APPLICANT: Koehn, Richard K.
;  APPLICANT:  Ruffner, Duane E.
;  APPLICANT:  Prakash, Ramesh K.
;  TITLE OF INVENTION: RNA-Based Inhibitory Oligonucleotides
;  FILE REFERENCE: 3302.2.7
;  CURRENT APPLICATION NUMBER: US/10/273,678
;  CURRENT FILING DATE:  2002-10-18
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 8705
;   TYPE: DNA
;   ORGANISM: Artificial
;   FEATURE:
;   OTHER INFORMATION: pShuttle
US-10-273-678-16

  Query Match             1.1%;  Score 70.8;  DB 6;  Length 8705;
  Best Local Similarity   45.3%;  Pred. No. 1.6e-06;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy    946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db   7954 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 7895

Qy   1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db   7894 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 7835

Qy   1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db   7834 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 7775

Qy   1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db   7774 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 7715

Qy   1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db   7714 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 7655

Qy   1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db   7654 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 7596

Qy   1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db   7595 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 7538

Qy   1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db   7537 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 7478

Qy   1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db   7477 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 7418

Qy   1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db   7417 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 7358

Qy   1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
```

```
Db   7357 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 7298

Qy   1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665

Db   7297 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 7238

Qy   1666 CTGATGAGCAGAAGGGAGAG 1685

Db   7237 TGGAGGCCGCCGGGGTAGAG 7218
```

Search completed: December 13, 2002, 00:06:21
Job time : 448 secs

GenCore version 5.1.3

OM protein - protein search, using sw model

Run on: December 13, 2002, 00:12:21 ; Search time 329 Seconds (without alignments) 3127.641 Million cell updates/sec

Title: US-09-902-432-4
Perfect score: 8073
Sequence: 1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 4569144

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/2/paa/PCTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8073	100.0	1596	1	PCT-US97-06830-4	Sequence 4, Appli
2	8073	100.0	1596	10	US-08-665-401-4	Sequence 4, Appli
3	8073	100.0	1596	13	US-08-978-277A-4	Sequence 4, Appli
4	8073	100.0	1596	23	US-09-902-432-4	Sequence 4, Appli
5	5421.5	67.2	1346	1	PCT-US97-06830-2	Sequence 2, Appli
6	5421.5	67.2	1346	10	US-08-665-401-2	Sequence 2, Appli
7	5421.5	67.2	1346	13	US-08-978-277A-2	Sequence 2, Appli
8	5421.5	67.2	1346	23	US-09-902-432-2	Sequence 2, Appli
9	4318.5	53.5	1782	27	US-60-389-987-2097	Sequence 2097, Ap
10	4318.5	53.5	1782	27	US-60-412-418-2097	Sequence 2097, Ap
11	4294	53.2	1781	21	US-09-738-877-3	Sequence 3, Appli
12	4294	53.2	1781	23	US-09-961-403-13	Sequence 13, Appl
13	4262.5	52.8	1795	1	PCT-US01-08631-51377	Sequence 51377, A
14	4253	52.7	1779	27	US-60-245-228-307	Sequence 307, App
15	780	9.7	732	21	US-09-758-442-562	Sequence 562, App
16	780	9.7	732	26	US-10-217-623-562	Sequence 562, App
17	752	9.3	212	27	US-60-182-569-1312	Sequence 1312, Ap
18	642	8.0	2768	20	US-09-614-150-31983	Sequence 31983, A
19	642	8.0	2768	27	US-60-173-464-24710	Sequence 24710, A
20	642	8.0	2768	27	US-60-191-637-31555	Sequence 31555, A
21	642	8.0	2768	27	US-60-191-681-25036	Sequence 25036, A
22	610.5	7.6	174	27	US-60-182-093-2249	Sequence 2249, Ap
23	561	6.9	194	27	US-60-186-662-1015	Sequence 1015, Ap
24	557.5	6.9	6358	27	US-60-191-637-27445	Sequence 27445, A
25	557.5	6.9	6358	27	US-60-191-681-22054	Sequence 22054, A
26	557.5	6.9	6815	20	US-09-614-150-27225	Sequence 27225, A
27	541.5	6.7	170	27	US-60-196-710-5536	Sequence 5536, Ap
28	536	6.6	144	27	US-60-178-307-2458	Sequence 2458, Ap
29	534.5	6.6	169	27	US-60-196-710-6834	Sequence 6834, Ap
30	493	6.1	3924	19	US-09-538-092-1246	Sequence 1246, Ap
31	492	6.1	2468	19	US-09-538-092-1135	Sequence 1135, Ap
32	492	6.1	2468	23	US-09-976-594-726	Sequence 726, App
33	492	6.1	2468	27	US-60-240-409-726	Sequence 726, App
34	492	6.1	2519	1	PCT-US01-08631-46995	Sequence 46995, A
35	489	6.1	6642	27	US-60-360-039-5013	Sequence 5013, Ap
36	485	6.0	3263	20	US-09-614-150-28422	Sequence 28422, A
37	485	6.0	3267	27	US-60-191-637-27442	Sequence 27442, A
38	485	6.0	3267	27	US-60-191-681-22051	Sequence 22051, A
39	484	6.0	178	27	US-60-188-162-4950	Sequence 4950, Ap
40	473.5	5.9	8991	11	US-08-714-741-32	Sequence 32, Appl
41	470	5.8	2364	26	US-10-205-331-66	Sequence 66, Appl
42	465.5	5.8	212	27	US-60-196-710-4143	Sequence 4143, Ap
43	464.5	5.8	2024	27	US-60-167-217-5735	Sequence 5735, Ap
44	464.5	5.8	2024	27	US-60-173-464-4625	Sequence 4625, Ap
45	462	5.7	1786	13	US-08-973-462-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1
PCT-US97-06830-4
; Sequence 4, Application PC/TUS9706830
; GENERAL INFORMATION:
; APPLICANT: Gelman, Irwin H.
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10112-0228
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US97/06830
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/665,401
; FILING DATE: 18-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Clark, Richard S
; REGISTRATION NUMBER: 26,154

```
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1596 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
PCT-US97-06830-4

  Query Match             100.0%;  Score 8073;  DB 1;  Length 1596;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1596;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60

Qy         61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120

Qy        121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy        181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240

Qy        241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300

Qy        301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360

Qy        361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420

Qy        421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480

Qy        481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540

Qy        541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600

Qy        601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Qy        661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720

Qy        721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780

Qy        781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840

Qy        841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Qy        901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Qy        961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Qy       1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080

Qy       1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140

Qy       1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200

Qy       1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260

Qy       1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320

Qy       1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380

Qy       1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440

Qy       1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500

Qy       1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560

Qy       1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
              ||||||||||||||||||||||||||||||||||||
Db       1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596


RESULT 2
US-08-665-401-4
; Sequence 4, Application US/08665401
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
```

```
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/665,401
;       FILING DATE:  18-JUN-1996
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Richard S
;       REGISTRATION NUMBER:  26,154
;       REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-408-2558
;       TELEFAX:  212-765-2519
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  4:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1596 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
US-08-665-401-4

  Query Match             100.0%;  Score 8073;  DB 10;  Length 1596;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1596;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60

Qy       61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120

Qy      121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy      181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240

Qy      241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300

Qy      301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360

Qy      361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420

Qy      421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480

Qy      481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540

Qy      541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600

Qy      601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Qy      661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720

Qy      721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780

Qy      781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840

Qy      841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Qy      901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Qy      961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Qy     1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080

Qy     1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140

Qy     1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200

Qy     1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260

Qy     1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320

Qy     1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380

Qy     1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440

Qy     1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500

Qy     1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560

Qy     1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
            ||||||||||||||||||||||||||||||||||||
Db     1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596

RESULT 3
```

```
US-08-978-277A-4
; Sequence 4, Application US/08978277A
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,277A
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1596 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
US-08-978-277A-4

  Query Match            100.0%;  Score 8073;  DB 13;  Length 1596;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 1596;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60

Qy       61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120

Qy      121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy      181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240

Qy      241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300

Qy      301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360

Qy      361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420

Qy      421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480

Qy      481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540

Qy      541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600

Qy      601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Qy      661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720

Qy      721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780

Qy      781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840

Qy      841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Qy      901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Qy      961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Qy     1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080

Qy     1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140

Qy     1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200

Qy     1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260

Qy     1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320

Qy     1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380

Qy     1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
```

```
Qy    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500

Qy    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560

Qy    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
           ||||||||||||||||||||||||||||||||||||
Db    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596


RESULT 4
US-09-902-432-4
; Sequence 4, Application US/09902432
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1596
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-09-902-432-4

  Query Match             100.0%;  Score 8073;  DB 23;  Length 1596;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1596;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60

Qy      61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120

Qy     121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy     181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240

Qy     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300

Qy     301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360

Qy     361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420

Qy     421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480

Qy     481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540

Qy     541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600

Qy     601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Qy     661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720

Qy     721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780

Qy     781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840

Qy     841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Qy     901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Qy     961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Qy    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080

Qy    1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140

Qy    1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200

Qy    1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260

Qy    1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320

Qy    1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380

Qy    1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440

Qy    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500

Qy    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560

Qy    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
```

```
              ||||||||||||||||||||||||||||||||||||
Db   1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596


RESULT 5
PCT-US97-06830-2
; Sequence 2, Application PC/TUS9706830
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/06830
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1346 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
PCT-US97-06830-2

  Query Match             67.2%;  Score 5421.5;  DB 1;  Length 1346;
  Best Local Similarity   91.6%;  Pred. No. 4.9e-304;
  Matches 1098;  Conservative   24;  Mismatches   62;  Indels   15;  Gaps    7;

Qy    387 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 60

Qy    447 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 120

Qy    507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 180

Qy    567 EKGPLEAPQDGEAEEGTTSDGEKKRE----GITPWASFKKMVTPKKRVRRPSESDKEEEL 622
          ||||||: |:  :         ||||:       ||          :|    :: ||||||||||||
Db    181 EKGPLEAPRMGKLRKELLR-GEKKRKDHSLGI-----LQKDGDTQETVRRPSESDKEEEL 234

Qy    623 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    235 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 294

Qy    683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 742
          ||||   |||||||   ::     : : :  |||||||||||||||||||||||||||||||
Db    295 SSSDIR-GPRTLGGGQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 353

Qy    743 VSTWESFKRLVTPRKKSKSKLEEK-AEDSSVEQLSTEIEPSREESWVS--IKKFIPGRRK 799
          ||||||||||||||||||||||| |  :  |      :  :| |  :    ::    |   :
Db    354 VSTWESFKRLVTPRKKSKSKLEEKEAGRTLVVGAGCPLRSNRVEKNLGFPLRNSSPDGGR 413

Qy    800 KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVY 859
          |   |:||||||||||||||:|:|||||||||:|||||||||||  |||    |  |
Db    414 KGQMGRQEQATVEDSGPVEINEDEPDVPAVVPLSEYDAVEREKMEAQGNAELPSCWGCV- 472

Qy    860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 532

Qy    920 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 592

Qy    980 DMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039
          ||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    593 DMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 652

Qy   1040 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 1099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    653 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 712

Qy   1100 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 1159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    713 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 772

Qy   1160 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELT 1219
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    773 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELA 832

Qy   1220 AAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
          |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    833 AAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 892

Qy   1280 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPS 1339
          ||||||||||||||||||||||||||||||||||||||||||||||   |||||||||||
Db    893 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVLTLDMPS 952

Qy   1340 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    953 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1012

Qy   1400 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1013 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1072

Qy   1460 ASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1519
          ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db   1073 ASQRERSEEEDKPDAGPDADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1132

Qy   1520 RTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAW 1578
          ||||||||||||||||| |||||||||||||| ||||||||||||||||||||||
Db   1133 RTETAPETHAYDSQTQVPAMRLDSREPNRCWTKMKVAKMKHPVPQPREDLQVLTVLEAW 1191


RESULT 6
US-08-665-401-2
; Sequence 2, Application US/08665401
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
```

```
;       CITY:  New York
;       STATE:  NY
;       COUNTRY:  USA
;       ZIP:  10112-0228
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Diskette
;       COMPUTER:  IBM Compatible
;       OPERATING SYSTEM:  DOS
;       SOFTWARE:  FastSEQ Version 1.5
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/665,401
;       FILING DATE:  18-JUN-1996
;       CLASSIFICATION:  514
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Clark, Richard S
;       REGISTRATION NUMBER:  26,154
;       REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-408-2558
;       TELEFAX:  212-765-2519
;       TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1346 amino acids
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  peptide
;     HYPOTHETICAL:  NO
;     ANTI-SENSE:  NO
;     FRAGMENT TYPE:  internal
;     ORIGINAL SOURCE:
US-08-665-401-2

  Query Match             67.2%;  Score 5421.5;  DB 10;  Length 1346;
  Best Local Similarity   91.6%;  Pred. No. 4.9e-304;
  Matches 1098;  Conservative   24;  Mismatches   62;  Indels   15;  Gaps    7;

Qy    387 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 446
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 60

Qy    447 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 120

Qy    507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 180

Qy    567 EKGPLEAPQDGEAEEGTTSDGEKKRE----GITPWASFKKMVTPKKRVRRPSESDKEEEL 622
          ||||||||: |:  :      ||||:    ||       :|      :: ||||||||||||
Db    181 EKGPLEAPRMGKLRKELLR-GEKKRKDHSLGI-----LQKDGDTQETVRRPSESDKEEEL 234

Qy    623 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 682
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    235 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 294

Qy    683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 742
          ||||  ||||||     ::      : : : ||||||||||||||||||||||||||||||
Db    295 SSSDIR-GPRTLGGGQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 353

Qy    743 VSTWESFKRLVTPRKKSKSKLEEK-AEDSSVEQLSTEIEPSREESWVS--IKKFIPGRRK 799
          ||||||||||||||||||||||||| |  : |       :  :| |  :   ::    |    :
Db    354 VSTWESFKRLVTPRKKSKSKLEEKEAGRTLVVGAGCPLRSNRVEKNLGFPLRNSSPDGGR 413

Qy    800 KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVY 859
          |  |:||||||||||||||||:|:|||||||||:||||||||||| |||    | |
Db    414 KGQMGRQEQATVEDSGPVEINEDEPDVPAVVPLSEYDAVEREKMEAQGNAELPSCWGCV- 472

Qy    860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 532

Qy    920 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 979
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 592

Qy    980 DMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039
          |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db    593 DMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 652

Qy   1040 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 1099
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    653 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 712

Qy   1100 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 1159
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    713 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 772

Qy   1160 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELT 1219
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||| 
Db    773 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELA 832

Qy   1220 AAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
          ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    833 AAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 892

Qy   1280 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPS 1339
          ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db    893 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVLTLDMPS 952

Qy   1340 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    953 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1012

Qy   1400 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1459
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1013 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1072

Qy   1460 ASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1519
          |||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db   1073 ASQRERSEEEDKPDAGPDADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1132

Qy   1520 RTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAW 1578
          ||||||||||||||||||| ||||||||||||||| |||||||||||||||||||||||
Db   1133 RTETAPETHAYDSQTQVPAMRLDSREPNRCWTKMKVAKMKHPVPQPREDLQVLTVLEAW 1191


RESULT 7
US-08-978-277A-2
; Sequence 2, Application US/08978277A
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,277A
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
```

```
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;   INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1346 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
US-08-978-277A-2

  Query Match             67.2%;  Score 5421.5;  DB 13;  Length 1346;
  Best Local Similarity   91.6%;  Pred. No. 4.9e-304;
  Matches 1098;  Conservative   24;  Mismatches   62;  Indels   15;  Gaps    7;

Qy        387 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 60

Qy        447 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 120

Qy        507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 180

Qy        567 EKGPLEAPQDGEAEEGTTSDGEKKRE----GITPWASFKKMVTPKKRVRRPSESDKEEEL 622
              |||||||: |:  :      ||||:     ||       :|  . :: ||||||||||||
Db        181 EKGPLEAPRMGKLRKELLR-GEKKRKDHSLGI-----LQKDGDTQETVRRPSESDKEEEL 234

Qy        623 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 294

Qy        683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 742
              ||||  |||||||   ::       : : : ||||||||||||||||||||||||||||||
Db        295 SSSDIR-GPRTLGGGQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 353

Qy        743 VSTWESFKRLVTPRKKSKSKLEEK-AEDSSVEQLSTEIEPSREESWVS--IKKFIPGRRK 799
              ||||||||||||||||||||||| |  : |       :  :| |  :  ::   |   :
Db        354 VSTWESFKRLVTPRKKSKSKLEEKEAGRTLVVGAGCPLRSNRVEKNLGFPLRNSSPDGGR 413

Qy        800 KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVY 859
              |    |:|||||||||||||||:|:||||||||||:||||||||||| |||     | |
Db        414 KGQMGRQEQATVEDSGPVEINEDEPDVPAVVPLSEYDAVEREKMEAQGNAELPSCWGCV- 472

Qy        860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 532

Qy        920 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 592

Qy        980 DMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039
              |||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db        593 DMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 652

Qy       1040 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 1099
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 712

Qy       1100 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 1159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 772

Qy       1160 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELT 1219
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELA 832

Qy       1220 AAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
              |||||  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        833 AAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 892

Qy       1280 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPS 1339
              ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db        893 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVLTLDMPS 952

Qy       1340 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399
              ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1012

Qy       1400 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1013 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1072

Qy       1460 ASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1519
              ||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||||
Db       1073 ASQRERSEEEDKPDAGPDADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1132

Qy       1520 RTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAW 1578
              ||||||||||||||||||  ||||||||||||||| ||||||||||||||||||||||
Db       1133 RTETAPETHAYDSQTQVPAMRLDSREPNRCWTKMKVAKMKHPVPQPREDLQVLTVLEAW 1191


RESULT 8
US-09-902-432-2
; Sequence 2, Application US/09902432
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-09-902-432-2

  Query Match             67.2%;  Score 5421.5;  DB 23;  Length 1346;
  Best Local Similarity   91.6%;  Pred. No. 4.9e-304;
  Matches 1098;  Conservative   24;  Mismatches   62;  Indels   15;  Gaps    7;

Qy        387 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 60

Qy        447 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 120

Qy        507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     121 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 180

Qy     567 EKGPLEAPQDGEAEEGTTSDGEKKRE----GITPWASFKKMVTPKKRVRRPSESDKEEEL 622
Db     181 EKGPLEAPRMGKLRKELLR-GEKKRKDHSLGI-----LQKDGDTQETVRRPSESDKEEEL 234

Qy     623 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 682
Db     235 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 294

Qy     683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 742
Db     295 SSSDIR-GPRTLGGGQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 353

Qy     743 VSTWESFKRLVTPRKKSKSKLEEK-AEDSSVEQLSTEIEPSREESWVS--IKKFIPGRRK 799
Db     354 VSTWESFKRLVTPRKKSKSKLEEKEAGRTLVVGAGCPLRSNRVEKNLGFPLRNSSPDGGR 413

Qy     800 KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVY 859
Db     414 KGQMGRQEQATVEDSGPVEINEDEPDVPAVVPLSEYDAVEREKMEAQGNAELPSCWGCV- 472

Qy     860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919
Db     473 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 532

Qy     920 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 979
Db     533 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 592

Qy     980 DMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039
Db     593 DMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 652

Qy    1040 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 1099
Db     653 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 712

Qy    1100 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 1159
Db     713 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 772

Qy    1160 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELT 1219
Db     773 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELA 832

Qy    1220 AAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
Db     833 AAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 892

Qy    1280 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPS 1339
Db     893 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVLTLDMPS 952

Qy    1340 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399
Db     953 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1012

Qy    1400 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1459
Db    1013 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1072

Qy    1460 ASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1519
Db    1073 ASQRERSEEEDKPDAGPDADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1132

Qy    1520 RTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAW 1578
Db    1133 RTETAPETHAYDSQTQVPAMRLDSREPNRCWTKMKVAKMKHPVPQPREDLQVLTVLEAW 1191
```

RESULT 9
US-60-389-987-2097

```
; Sequence 2097, Application US/60389987
; GENERAL INFORMATION:
;  APPLICANT: Ghosh, Soumitra S.
;  APPLICANT:  Fahy, Eoin D.
;  APPLICANT:  Zhang, Bing
;  APPLICANT:  Gibson, Bradford W.
;  APPLICANT:  Taylor, Steven W.
;  APPLICANT:  Glenn, Gary M.
;  APPLICANT:  Warnock, Dale E.
;  TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
;  TITLE OF INVENTION:  IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
;  FILE REFERENCE: 660088.465P2
;  CURRENT APPLICATION NUMBER: US/60/389,987
;  CURRENT FILING DATE:  2002-06-17
;  NUMBER OF SEQ ID NOS: 3025
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2097
;   LENGTH: 1782
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-60-389-987-2097

  Query Match             53.5%;  Score 4318.5;  DB 27;  Length 1782;
  Best Local Similarity   58.1%;  Pred. No. 4.7e-240;
  Matches  961;  Conservative  189;  Mismatches  398;  Indels  107;  Gaps   34;

Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db       1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy      57 QLSSVNGVAEQGDVHVQEENQEGQ-------------EEEVVDEDVGQRESEDVREKDR 102
Db      61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSKRDS 120

Qy     103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db     121 DKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 180

Qy     163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db     181 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHKDPSL--GAGEAASKESEPKQS 235

Qy     223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db     236 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 294

Qy     283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db     295 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 354

Qy     338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db     355 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 414

Qy     393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db     415 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDAEPQEAEPAKELVKL 466

Qy     453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db     467 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 526

Qy     513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db     527 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 585

Qy     573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db     586 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 645

Qy     633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db     646 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPK 705
```

```
Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db    706 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 765

Qy    753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
Db    766 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 825

Qy    809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db    826 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELS 885

Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db    886 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 945

Qy    925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db    946 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSA 1005

Qy    985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db   1006 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1065

Qy   1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db   1066 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1119

Qy   1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db   1120 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1179

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db   1180 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1238

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db   1239 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1296

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db   1297 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1356

Qy   1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db   1357 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1416

Qy   1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db   1417 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1476

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db   1477 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1533

Qy   1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547
Db   1534 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1591

Qy   1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db   1592 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1626
```

RESULT 10
US-60-412-418-2097
; Sequence 2097, Application US/60412418
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465P3
; CURRENT APPLICATION NUMBER: US/60/412,418
; CURRENT FILING DATE: 2002-09-20
; NUMBER OF SEQ ID NOS: 3025
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2097
; LENGTH: 1782
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-412-418-2097

```
  Query Match             53.5%;  Score 4318.5;  DB 27;  Length 1782;
  Best Local Similarity   58.1%;  Pred. No. 4.7e-240;
  Matches  961;  Conservative  189;  Mismatches  398;  Indels  107;  Gaps   34;

Qy      1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db      1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy     57 QLSSVNGVAEQGDVHVQEENQEGQ-------------EEEVVDEDVGQRESEDVREKDR 102
Db     61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSKRDS 120

Qy    103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db    121 DKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 180

Qy    163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db    181 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHKDPSL--GAGEAASKESEPKQS 235

Qy    223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db    236 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 294

Qy    283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db    295 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 354

Qy    338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db    355 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 414

Qy    393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db    415 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDAEPQEAEPAKELVKL 466

Qy    453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db    467 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 526

Qy    513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db    527 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 585

Qy    573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db    586 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 645

Qy    633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db    646 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPK 705

Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db    706 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 765
```

```
Qy     753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
Db     766 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 825

Qy     809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db     826 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELS 885

Qy     866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db     886 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 945

Qy     925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db     946 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSA 1005

Qy     985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db    1006 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1065

Qy    1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db    1066 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1119

Qy    1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db    1120 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1179

Qy    1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db    1180 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1238

Qy    1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db    1239 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1296

Qy    1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db    1297 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1356

Qy    1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db    1357 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1416

Qy    1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db    1417 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1476

Qy    1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db    1477 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1533

Qy    1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547
Db    1534 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1591

Qy    1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db    1592 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1626
```

RESULT 11
US-09-738-877-3
; Sequence 3, Application US/09738877
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan
; APPLICANT: Weiss, Stephen J.
; APPLICANT: Glynne, Richard
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET
; TITLE OF INVENTION: SCREENING FOR ANGIOGENESIS MODULATORS
; FILE REFERENCE: A-69806/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/738,877
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,425
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT/US 00/22061
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
;   LENGTH: 1781
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-738-877-3

```
  Query Match             53.2%;  Score 4294;  DB 21;  Length 1781;
  Best Local Similarity   57.9%;  Pred. No. 1.2e-238;
  Matches  959;  Conservative  187;  Mismatches  401;  Indels  108;  Gaps   35;

Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db       1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy      57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102
Db      61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy     103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db     121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy     163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db     180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy     223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db     235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy     283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db     294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy     338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db     354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy     393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db     414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy     453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db     466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy     513 KLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db     526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy     573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db     585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy     633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db     645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPK 704

Qy     693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db     705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy     753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
```

```
Db     765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy     809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865

Db     825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy     866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924

Db     885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy     925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984

Db     945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy     985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038

Db    1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy    1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097

Db    1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy    1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147

Db    1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy    1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207

Db    1179 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1237

Qy    1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259

Db    1238 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1295

Qy    1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314

Db    1296 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1355

Qy    1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372

Db    1356 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1415

Qy    1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430

Db    1416 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1475

Qy    1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEDKPDAGPDADGKESTAIEKVL 1490

Db    1476 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1532

Qy    1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547

Db    1533 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1590

Qy    1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577

Db    1591 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1625
```

RESULT 12
US-09-961-403-13
; Sequence 13, Application US/09961403
; GENERAL INFORMATION:
; APPLICANT: HE-STUMPP, HOLGER
; APPLICANT: HAENDLER, BERNARD
; APPLICANT: KRAETZSCHMAR, JOERN
; APPLICANT: KREFT, BERTHOLT
; APPLICANT: WINTERHAGER, ELKE
; APPLICANT: REGIDOR, PEDRO
; APPLICANT: SCOTTI, SIMONE
; TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
;   LENGTH: 1781
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-961-403-13

```
  Query Match             53.2%;  Score 4294;  DB 23;  Length 1781;
  Best Local Similarity   57.9%;  Pred. No. 1.2e-238;
  Matches  959;  Conservative  187;  Mismatches  401;  Indels  108;  Gaps   35;

Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56

Db       1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy      57 QLSSVNGVAEQGDVHVQEENQEGQ-------------EEEVVDEDVGQRESEDVREKDR 102

Db      61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy     103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162

Db     121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy     163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222

Db     180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy     223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282

Db     235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy     283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337

Db     294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy     338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392

Db     354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy     393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452

Db     414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy     453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512

Db     466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy     513 KLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572

Db     526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy     573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632

Db     585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy     633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692

Db     645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRSSSDEEGGPK 704

Qy     693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752

Db     705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy     753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808

Db     765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy     809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865

Db     825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884
```

```
Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924

Db    885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy    925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984

Db    945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy    985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038

Db   1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy   1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097

Db   1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy   1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147

Db   1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207

Db   1179 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1237

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259

Db   1238 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1295

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314

Db   1296 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1355

Qy   1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372

Db   1356 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1415

Qy   1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430

Db   1416 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1475

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490

Db   1476 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1532

Qy   1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547

Db   1533 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1590

Qy   1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577

Db   1591 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1625
```

RESULT 13
PCT-US01-08631-51377
; Sequence 51377, Application PC/TUS0108631
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc
; TITLE OF INVENTION: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES
; FILE REFERENCE: 21272-049
; CURRENT APPLICATION NUMBER: PCT/US01/08631
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 09/540,217
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 09/649,167
; PRIOR FILING DATE: 2000-08-23
; NUMBER OF SEQ ID NOS: 60736
; SOFTWARE: Custom
; SEQ ID NO 51377
;   LENGTH: 1795
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: DOMAIN
;   LOCATION: (415)..(463)
;   OTHER INFORMATION: G-protein gamma subunit profile domain identified by eMATRIX,
;   OTHER INFORMATION: accession number BL50058, p-value=5.557e-09, raw score of 27.23
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1795)
;   OTHER INFORMATION: Xaa = X or * as defined in Table 2
PCT-US01-08631-51377

```
  Query Match             52.8%;  Score 4262.5;  DB 1;  Length 1795;
  Best Local Similarity   57.5%;  Pred. No. 8.2e-237;
  Matches  951;  Conservative  190;  Mismatches  410;  Indels  103;  Gaps   33;

Qy      1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56

Db      9 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 68

Qy     57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102

Db     69 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVRQVDHIDSEDVSDRDS 128

Qy    103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162

Db    129 DTEMTTKSAVVHDITDDGQEETPDIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 188

Qy    163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222

Db    189 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 243

Qy    223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282

Db    244 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 302

Qy    283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337

Db    303 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 362

Qy    338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392

Db    363 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 422

Qy    393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452

Db    423 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 474

Qy    453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512

Db    475 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 534

Qy    513 KLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572

Db    535 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 593

Qy    573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632

Db    594 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 653

Qy    633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692

Db    654 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPK 713

Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752

Db    714 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 773

Qy    753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808

Db    774 VTPGKKSKSKLEEKSEDSLAGSGVEHSTPDTEPGKEESWVSIKKFIPGLRKKRPDGKQEQ 833

Qy    809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865

Db    834 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELS 893
```

```
Qy     866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db     894 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 953

Qy     925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db     954 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSA 1013

Qy     985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db    1014 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1073

Qy    1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db    1074 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1127

Qy    1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db    1128 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1187

Qy    1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db    1188 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1246

Qy    1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db    1247 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1304

Qy    1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETK-PEQVSE 1313
Db    1305 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREEYQXQSPTHVNE 1364

Qy    1314 EG-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSL 1371
Db    1365 EKLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSS 1424

Qy    1372 DTTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPL 1429
Db    1425 EASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPT 1484

Qy    1430 GPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKV 1489
Db    1485 GPDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEED 1542

Qy    1490 LKAEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNR 1548
Db    1543 LEPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQ 1602

Qy    1549 CWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db    1603 ETEKEGEDPQASAQDETPITSAKEESESTAVGQA 1636
```

RESULT 14
US-60-245-228-307
; Sequence 307, Application US/60245228
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN CYCLASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN CYCLASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000878
; CURRENT APPLICATION NUMBER: US/60/245,228
; CURRENT FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 630
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 307
; LENGTH: 1779
; TYPE: PRT
; ORGANISM: HUMAN
US-60-245-228-307

Query Match 52.7%; Score 4253; DB 27; Length 1779;
Best Local Similarity 57.6%; Pred. No. 2.9e-236;
Matches 952; Conservative 184; Mismatches 406; Indels 110; Gaps 33;

```
Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEASGAAGDPADA------DPATKLPQ 53
Db       9 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPWAE---AAPDPPAAPTIAAWDPATKLLQ 65

Qy      54 KNGQLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVRE 99
Db      66 KNGQLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSD 125

Qy     100 KDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFK 159
Db     126 SDK--EMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQANDIGFKKVFK 183

Qy     160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESEL 219
Db     184 FVGFKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEP 238

Qy     220 KQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSE 279
Db     239 KQSTEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSE 297

Qy     280 TTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE-- 337
Db     298 TGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLT 357

Qy     338 -------QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAH 390
Db     358 ASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVH 417

Qy     391 Q-EVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEEL 449
Db     418 QEEVVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDAEPQEAEPAKEL 469

Qy     450 MKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSS 509
Db     470 VKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTST 529

Qy     510 GLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKG 569
Db     530 GLKKLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKG 588

Qy     570 PLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSAT 629
Db     589 LAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSAT 648

Qy     630 LSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEG 689
Db     649 LSSTESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEG 708

Qy     690 GPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESF 749
Db     709 GPKAMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESF 768

Qy     750 KRLVTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGK 805
Db     769 KRLVTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGK 828

Qy     806 QEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSE 862
Db     829 QEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSK 888

Qy     863 ELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDII 922
Db     889 ELSESQVHMMAAAVADGTRAATIIEERSPSWISASVTGPLEQVEAEAALLTEEVLEREVI 948

Qy     923 A-EETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDM 981
Db     949 AEEEPPTVTEPLPENRER--DTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEM 1006

Qy     982 VSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT--- 1038
```

```
Db    1007 VSAVSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGP 1066

Qy    1039 ----QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQAT 1094
               | |||      : |  || :|    :|| ||: |   |||   |   ||    || |
Db    1067 EDVLQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTT 1120

Qy    1095 PESLE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTP 1144
           ||| |  |:||  ::     |||       :| |:: ||||: |:| || |||||:||||
Db    1121 PESFEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTP 1180

Qy    1145 LADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGE 1204
           :|| |     ||:|| ::  :      |:  ||  || ||  |||| |     |::   ||| |
Db    1181 VADFDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKE 1239

Qy    1205 EPG-RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSGPN 1258
           :    | || | :|::    | :|:|||  |   | |      || |      |  |  : :|
Db    1240 QSKMEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGIT 1297

Qy    1259 SQKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSE 1313
            :             |      | |:: :: |:||     |  | ||.  ||:|| | :|  |:|
Db    1298 VSREKVTEVKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNE 1357

Qy    1314 EG-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSL 1371
             |  | |||      |      : : |:::|  :   | : || ||  |      |::| |  ::||||
Db    1358 EKLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSS 1417

Qy    1372 DTTVTQTAEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGP 1431
           : : | || | |: :       : |||: |   ||||:  ||||        :      || || ||
Db    1418 EASFTLTAAAEEEKVLGETANITGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGP 1477

Qy    1432 ESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLK 1491
           : ||:| |:||:    :   |   ||:|| : | : :|:| |:  |   :     | | |||: |
Db    1478 DCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEEDL- 1534

Qy    1492 AEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRCW 1550
            |||     ||:||:|:| |:|||||||| || ||| |       : |||      : ||:: :
Db    1535 -EPENGILETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQET 1593

Qy    1551 TK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
            |        |: : |:   :|: :     | :|
Db    1594 EKEGEEPQASAQDETPITSAKEESESTAVGQA 1625
```

RESULT 15
US-09-758-442-562
; Sequence 562, Application US/09758442
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PM031
; CURRENT APPLICATION NUMBER: US/09/758,442
; CURRENT FILING DATE: 2001-01-11
; PRIOR APPLICATION NUMBER: 60/179,065
; PRIOR FILING DATE: 2000-01-31
; PRIOR APPLICATION NUMBER: 60/180,628
; PRIOR FILING DATE: 2000-02-04
; NUMBER OF SEQ ID NOS: 710
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 562
;   LENGTH: 732
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (67)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (162)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (164)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
;   NAME/KEY: SITE
;   LOCATION: (464)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-758-442-562

```
  Query Match             9.7%;  Score 780;  DB 21;  Length 732;
  Best Local Similarity   38.6%;  Pred. No. 5.8e-36;
  Matches  221;  Conservative   87;  Mismatches  220;  Indels   44;  Gaps   17;

Qy    1039 QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESL 1098
           | |||      : |  || :|    :|| ||: |   |||   |   ||    || ||||
Db       3 QPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPESF 56

Qy    1099 E-VPEVTADV---DHVATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADS 1148
           |  |:||  :    : | |||       :| |:: ||||: |:| || |||||:||||:||
Db      57 EKAPQVTESIXSSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADF 116

Qy    1149 DTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG- 1207
           |     ||:|| ::  :      |:  ||  || ||  |||| |     |::   ||| |:
Db     117 DAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSXFXFQEETKEQSKM 175

Qy    1208 RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQ 1260
            | || | :|::    | :|:|||  |   | |      || |      |  |  : :|    :
Db     176 EDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSRE 233

Qy    1261 KAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG 1315
           |  :|     |      | |:: :: |:||     |  | ||   ||:|| | :|  |:||
Db     234 KVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEK 293

Qy    1316 -EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDT 1373
             | |||      |      : : |:::|  :   | : || ||  |      |::| |  ::|||| :
Db     294 LEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEA 353

Qy    1374 TVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGP 1431
           : | || | |: :      ||  | ||||:  ||||        :      || || ||
Db     354 SFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGP 413

Qy    1432 ESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLK 1491
           : ||:| |:||:    :   |   ||:|| : | : :|:| |:  |   :     | | |||: |:
Db     414 DCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSXAIEEDLE 471

Qy    1492 AEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRCW 1550
            | |||||:||:|:| |:|||||||| || ||| |       : |||      : ||:: :
Db     472 PENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQET 531

Qy    1551 TK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
            |        |: : |:   :|: :     | :|
Db     532 EKEGEDPQASAQDETPITSAKEESESTAVGQA 563
```

Search completed: December 13, 2002, 00:22:46
Job time : 339 secs

```
                    GenCore version 5.1.3
           Copyright (c) 1993 - 2002  Compugen Ltd.

OM protein - protein search, using sw model

Run on:         December 13, 2002, 00:13:41 ; Search time 85 Seconds
                                          (without alignments)
                                          1253.621 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       237906 seqs, 66765526 residues

Total number of hits satisfying chosen parameters:      237906

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Pending_Patents_AA_New:*
                1:  /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep:*
                2:  /cgn2_6/ptodata/2/paa/US06_NEW_COMB.pep:*
                3:  /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep:*
                4:  /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep:*
                5:  /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep:*
                6:  /cgn2_6/ptodata/2/paa/US10_NEW_COMB.pep:*
                7:  /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4335	53.7	1844	5	US-09-724-676-61346	Sequence 61346, A
2	4335	53.7	1844	5	US-09-724-676A-61346	Sequence 61346, A
3	4294	53.2	1781	1	PCT-US02-04915-44	Sequence 44, Appl
4	4098	50.8	1685	5	US-09-724-676-61347	Sequence 61347, A
5	4098	50.8	1685	5	US-09-724-676A-61347	Sequence 61347, A
6	429	5.3	2665	6	US-10-203-138A-11282	Sequence 11282, A
7	401.5	5.0	4368	5	US-09-724-676-73643	Sequence 73643, A
8	401.5	5.0	4368	5	US-09-724-676A-73643	Sequence 73643, A
9	400	5.0	2642	5	US-09-724-676-55851	Sequence 55851, A
10	400	5.0	2642	5	US-09-724-676A-55851	Sequence 55851, A
11	400	5.0	2722	5	US-09-724-676-55868	Sequence 55868, A
12	400	5.0	2722	5	US-09-724-676A-55868	Sequence 55868, A
13	400	5.0	2915	5	US-09-724-676-55849	Sequence 55849, A
14	400	5.0	2915	5	US-09-724-676-55850	Sequence 55850, A
15	400	5.0	2915	5	US-09-724-676A-55849	Sequence 55849, A
16	400	5.0	2915	5	US-09-724-676A-55850	Sequence 55850, A
17	400	5.0	2929	5	US-09-724-676-55846	Sequence 55846, A
18	400	5.0	2929	5	US-09-724-676A-55846	Sequence 55846, A
19	400	5.0	2940	5	US-09-724-676-55847	Sequence 55847, A
20	400	5.0	2940	5	US-09-724-676-55848	Sequence 55848, A
21	400	5.0	2940	5	US-09-724-676A-55847	Sequence 55847, A
22	400	5.0	2940	5	US-09-724-676A-55848	Sequence 55848, A
23	400	5.0	2995	5	US-09-724-676-55866	Sequence 55866, A
24	400	5.0	2995	5	US-09-724-676-55867	Sequence 55867, A
25	400	5.0	2995	5	US-09-724-676A-55866	Sequence 55866, A
26	400	5.0	2995	5	US-09-724-676A-55867	Sequence 55867, A
27	400	5.0	3009	5	US-09-724-676-55863	Sequence 55863, A
28	400	5.0	3009	5	US-09-724-676A-55863	Sequence 55863, A
29	400	5.0	3020	5	US-09-724-676-55864	Sequence 55864, A
30	400	5.0	3020	5	US-09-724-676-55865	Sequence 55865, A
31	400	5.0	3020	5	US-09-724-676A-55864	Sequence 55864, A
32	400	5.0	3020	5	US-09-724-676A-55865	Sequence 55865, A
33	388	4.8	1325	6	US-10-203-138A-12693	Sequence 12693, A
34	382	4.7	2014	5	US-09-724-676-49425	Sequence 49425, A
35	382	4.7	2014	5	US-09-724-676A-49425	Sequence 49425, A
36	380.5	4.7	2101	5	US-09-724-676-49404	Sequence 49404, A
37	380.5	4.7	2101	5	US-09-724-676-49405	Sequence 49405, A
38	380.5	4.7	2101	5	US-09-724-676-49419	Sequence 49419, A
39	380.5	4.7	2101	5	US-09-724-676-49420	Sequence 49420, A
40	380.5	4.7	2101	5	US-09-724-676-49421	Sequence 49421, A
41	380.5	4.7	2101	5	US-09-724-676-49422	Sequence 49422, A
42	380.5	4.7	2101	5	US-09-724-676-49423	Sequence 49423, A
43	380.5	4.7	2101	5	US-09-724-676-49424	Sequence 49424, A
44	380.5	4.7	2101	5	US-09-724-676A-49404	Sequence 49404, A
45	380.5	4.7	2101	5	US-09-724-676A-49405	Sequence 49405, A

ALIGNMENTS

```
RESULT 1
US-09-724-676-61346
; Sequence 61346, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 61346
;   LENGTH: 1844
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676-61346

  Query Match             53.7%;  Score 4335;  DB 5;  Length 1844;
  Best Local Similarity   58.2%;  Pred. No. 2.1e-248;
  Matches  962;  Conservative  188;  Mismatches  401;  Indels  102;  Gaps   32;

Qy          1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
              |||||||||||||| | || ||:|    || ||:|||:        ||| |  ||||| ||||
Db         62 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 121

Qy         57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102
              |||::||||| :: :|| :  ||              ||||:  :|||:|||| |:|
Db        122 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 181

Qy        103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
               :|||  |  | ||| |||||| ||||||:||:|:||: || ||||||:||||||||||
Db        182 DKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 241

Qy        163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
              |||||||| || ||||||||||||||      ||||||:||:     ||:|||||| |||
Db        242 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 296

Qy        223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
              ||| | |||:|||  ||   ||| ||  || |:||||||||||:|| |||:||| ||| |
Db        297 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 355

Qy        283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
              :||||| ||||||||||||:| |||::| :||:||| ||||  ||    | | |||:
Db        356 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 415

Qy        338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
                  |||||  :  ||||:||||||| |:||   :  ||||  ||||||||||:| || |
Db        416 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 475
```

```
Qy    393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db    476 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDAEPQEAEPAKELVKL 527

Qy    453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db    528 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 587

Qy    513 KLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db    588 KLSGKKQKGKR-GGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 646

Qy    573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db    647 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 706

Qy    633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db    707 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPK 766

Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db    767 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 826

Qy    753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSRREESWVSIKKFIPGRRKKRADGKQEQ 808
Db    827 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 886

Qy    809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db    887 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELS 946

Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db    947 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 1006

Qy    925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db   1007 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSA 1066

Qy    985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db   1067 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1126

Qy   1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db   1127 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1180

Qy   1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db   1181 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1240

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db   1241 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1299

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db   1300 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1357

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db   1358 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1417

Qy   1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db   1418 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1477

Qy   1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db   1478 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1537

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db   1538 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEEDL 1595

Qy   1491 KAEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRC 1549
Db   1596 EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQE 1655

Qy   1550 WTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db   1656 TEKEGEEPQASAQDETPITSAKEESESTAVGQA 1688
```

RESULT 2
US-09-724-676A-61346
; Sequence 61346, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 61346
;   LENGTH: 1844
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676A-61346

```
  Query Match             53.7%;  Score 4335;  DB 5;  Length 1844;
  Best Local Similarity   58.2%;  Pred. No. 2.1e-248;
  Matches  962;  Conservative  188;  Mismatches  401;  Indels  102;  Gaps  32;

Qy      1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db     62 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 121

Qy     57 QLSSVNGVAEQGDVHVQEENQEGQ-------------EEEVVDEDVGQRESEDVREKDR 102
Db    122 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 181

Qy    103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db    182 DKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 241

Qy    163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db    242 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 296

Qy    223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db    297 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 355

Qy    283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db    356 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 415

Qy    338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db    416 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 475

Qy    393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db    476 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDAEPQEAEPAKELVKL 527

Qy    453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db    528 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 587

Qy    513 KLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db    588 KLSGKKQKGKR-GGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 646
```

```
Qy    573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db    647 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 706

Qy    633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db    707 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPK 766

Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db    767 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 826

Qy    753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
Db    827 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 886

Qy    809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db    887 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELS 946

Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db    947 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 1006

Qy    925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db   1007 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSA 1066

Qy    985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db   1067 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1126

Qy   1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db   1127 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1180

Qy   1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db   1181 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1240

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db   1241 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1299

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db   1300 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1357

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db   1358 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1417

Qy   1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db   1418 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1477

Qy   1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db   1478 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1537

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db   1538 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEEDL 1595

Qy   1491 KAEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRC 1549
Db   1596 EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQE 1655

Qy   1550 WTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db   1656 TEKEGEEPQASAQDETPITSAKEESESTAVGQA 1688
```

```
RESULT 3
PCT-US02-04915-44
; Sequence 44, Application PC/TUS0204915
; GENERAL INFORMATION:
;  APPLICANT: Murray, Richard
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Angiogenesis Modulators
;  FILE REFERENCE: 018501-006200PC
;  CURRENT APPLICATION NUMBER: PCT/US02/04915
;  CURRENT FILING DATE:  2002-02-14
;  PRIOR APPLICATION NUMBER: US 09/784,356
;  PRIOR FILING DATE: 2001-02-14
;  PRIOR APPLICATION NUMBER: US 09/791,390
;  PRIOR FILING DATE: 2001-02-22
;  PRIOR APPLICATION NUMBER: US 60/285,475
;  PRIOR FILING DATE: 2002-04-19
;  PRIOR APPLICATION NUMBER: US 60/310,025
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/334,244
;  PRIOR FILING DATE: 2001-11-29
;  NUMBER OF SEQ ID NOS: 230
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
;   LENGTH: 1781
;   TYPE: PRT
;   ORGANISM: Homo sapiens
PCT-US02-04915-44

  Query Match             53.2%;  Score 4294;  DB 1;  Length 1781;
  Best Local Similarity   57.9%;  Pred. No. 5.4e-246;
  Matches  959;  Conservative  187;  Mismatches  401;  Indels  108;  Gaps   35;

Qy      1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db      1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy     57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102
Db     61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy    103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db    121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy    163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db    180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy    223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db    235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy    283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db    294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy    338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db    354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy    393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db    414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy    453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db    466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525
```

```
Qy    513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572

Db    526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy    573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632

Db    585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy    633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692

Db    645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRSSSDEEGGPK 704

Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752

Db    705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy    753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808

Db    765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy    809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865

Db    825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924

Db    885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy    925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984

Db    945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy    985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038

Db   1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy   1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097

Db   1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy   1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147

Db   1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207

Db   1179 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1237

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259

Db   1238 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1295

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314

Db   1296 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1355

Qy   1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372

Db   1356 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1415

Qy   1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430

Db   1416 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1475

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490

Db   1476 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1532

Qy   1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547

Db   1533 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1590

Qy   1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577

Db   1591 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1625
```

RESULT 4
US-09-724-676-61347
; Sequence 61347, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 61347
;   LENGTH: 1685
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676-61347

```
  Query Match             50.8%;  Score 4098;  DB 5;  Length 1685;
  Best Local Similarity   58.4%;  Pred. No. 2.1e-234;
  Matches  904;  Conservative  177;  Mismatches  382;  Indels   84;  Gaps   28;

Qy     89 VGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQ 148

Db      9 VGQRDSEDVSERDSDKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQ 68

Qy    149 ANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETA 208

Db     69 ANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--G 123

Qy    209 VGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKS 268

Db    124 AGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKS 182

Qy    269 PESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEK 328

Db    183 AESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEED 242

Qy    329 EKTEPASEE---------QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLA 379

Db    243 GKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLA 302

Qy    380 TEVFDEKMEAHQ-EVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQE 438

Db    303 TEVFDEKIEVHQEEVVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDA 354

Qy    439 VPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQ 498

Db    355 EPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQ 414

Qy    499 GSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPE 558

Db    415 GSPLKKLFTSTGLKKLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPE 473

Qy    559 EPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDK 618

Db    474 EPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDK 533

Qy    619 EEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKR 678

Db    534 EDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKR 593

Qy    679 ARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPS 738

Db    594 ARRGSSSDEEGGPKAMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPT 653

Qy    739 EGEGVSTWESFKRLVTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFI 794

Db    654 EGEGVSTWESFKRLVTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFI 713
```

```
Qy     795 PGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTEL 851
Db     714 PGRRKKRPDGKQEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQ 773

Qy     852 PQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMP 911
Db     774 PEQKAATEVSKELSESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAAL 833

Qy     912 PVEEVTEKDIIA-EETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVT 970
Db     834 LTEEVLEREVIAEEEPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGT 893

Qy     971 EASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK 1030
Db     894 EASAAEETTEMVSAVSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVK 953

Qy    1031 EESQVPAT-------QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEH 1083
Db     954 EESQLPGTGGPEDVLQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEP 1007

Qy    1084 LAQGSETGQATPESLE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSET 1133
Db    1008 FTQGKVVGQTTPESFEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVET 1067

Qy    1134 LTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLP 1193
Db    1068 PTDSETDGSTPVADFDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-P 1126

Qy    1194 NNVPAQEEHGEEPG-RDVLEPTQQELTAAVPVLAKTEVGQEGEVDWLDGEKVK-----E 1247
Db    1127 SSFVFQEETKEQSKMEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFE 1184

Qy    1248 EQEVFVHSG--PNSQKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVE 1300
Db    1185 GLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVE 1244

Qy    1301 KEKRETKPEQVSEEG-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQ 1358
Db    1245 REKTEAEPTHVNEEKLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQ 1304

Qy    1359 DKAGCIEVQVQSLDTTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGG 1416
Db    1305 EEAVCTKIQVQSSEASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNE 1364

Qy    1417 HWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGP 1476
Db    1365 DFAAHPGEDAVPTGPDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQ 1424

Qy    1477 DADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQ 1535
Db    1425 EV--KVSVAIEEDLEPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQ 1482

Qy    1536 VPACRLDSREPNRCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db    1483 AHVIKADSQDAGQETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1529
```

RESULT 5
US-09-724-676A-61347
; Sequence 61347, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 61347
; LENGTH: 1685
; TYPE: PRT
; ORGANISM: Homo sapiens

US-09-724-676A-61347

```
  Query Match             50.8%;  Score 4098;  DB 5;  Length 1685;
  Best Local Similarity   58.4%;  Pred. No. 2.1e-234;
  Matches  904;  Conservative  177;  Mismatches  382;  Indels   84;  Gaps   28;

Qy      89 VGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQ 148
Db       9 VGQRDSEDVSERDSDKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQ 68

Qy     149 ANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETA 208
Db      69 ANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--G 123

Qy     209 VGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKS 268
Db     124 AGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKS 182

Qy     269 PESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEK 328
Db     183 AESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEED 242

Qy     329 EKTEPASEE---------QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLA 379
Db     243 GKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLA 302

Qy     380 TEVFDEKMEAHQ-EVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQE 438
Db     303 TEVFDEKIEVHQEEVVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDA 354

Qy     439 VPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQ 498
Db     355 EPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQ 414

Qy     499 GSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPE 558
Db     415 GSPLKKLFTSTGLKKLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPE 473

Qy     559 EPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDK 618
Db     474 EPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDK 533

Qy     619 EEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKR 678
Db     534 EDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKR 593

Qy     679 ARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSSPEPAGSPS 738
Db     594 ARRGSSSDEEGGPKAMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPT 653

Qy     739 EGEGVSTWESFKRLVTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFI 794
Db     654 EGEGVSTWESFKRLVTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFI 713

Qy     795 PGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTEL 851
Db     714 PGRRKKRPDGKQEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQ 773

Qy     852 PQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMP 911
Db     774 PEQKAATEVSKELSESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAAL 833

Qy     912 PVEEVTEKDIIA-EETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVT 970
Db     834 LTEEVLEREVIAEEEPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGT 893

Qy     971 EASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK 1030
Db     894 EASAAEETTEMVSAVSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVK 953

Qy    1031 EESQVPAT-------QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEH 1083
Db     954 EESQLPGTGGPEDVLQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEP 1007
```

```
Qy   1084 LAQGSETGQATPESLE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSET 1133
Db   1008 FTQGKVVGQTTPESFEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVET 1067

Qy   1134 LTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLP 1193
Db   1068 PTDSETDGSTPVADFDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-P 1126

Qy   1194 NNVPAQEEHGEEPG-RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----E 1247
Db   1127 SSFVFQEETKEQSKMEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFE 1184

Qy   1248 EQEVFVHSG--PNSQKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVE 1300
Db   1185 GLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVE 1244

Qy   1301 KEKRETKPEQVSEEG-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQ 1358
Db   1245 REKTEAEPTHVNEEKLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQ 1304

Qy   1359 DKAGCIEVQVQSLDTTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGG 1416
Db   1305 EEAVCTKIQVQSSEASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNE 1364

Qy   1417 HWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGP 1476
Db   1365 DFAAHPGEDAVPTGPDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQ 1424

Qy   1477 DADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQ 1535
Db   1425 EV--KVSVAIEEDLEPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQ 1482

Qy   1536 VPACRLDSREPNRCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db   1483 AHVIKADSQDAGQETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1529
```

RESULT 6
US-10-203-138A-11282
; Sequence 11282, Application US/10203138A
; GENERAL INFORMATION:
; APPLICANT: Molecular Dynamics, Inc.
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: ANALYSIS OF GENE EXPRESSION IN BT 474
; FILE REFERENCE: PB 0004 WO 8
; CURRENT APPLICATION NUMBER: US/10/203,138A
; CURRENT FILING DATE: 2002-08-02
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 04 February 2000 (04.02.00)
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 26 May 2000 (26.05.00)
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 03 August 2000 (03.08.00)
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 03 October 2000 (03.10.00)
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 27 September 2000 (27.09.00)
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 21 September 2000 (21.09.00)
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 30 June 2000 (30.06.00)
; NUMBER OF SEQ ID NOS: 15438
; SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 11282
; LENGTH: 2665
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL034555.2
; FEATURE:
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
; FEATURE:
; OTHER INFORMATION: EST_HUMAN HIT: AU117052.1, EVALUE 0.00e+00
; FEATURE:
; OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUE 3.00e-10
US-10-203-138A-11282

Query Match 5.3%; Score 429; DB 6; Length 2665;
Best Local Similarity 19.6%; Pred. No. 1.3e-16;
Matches 372; Conservative 270; Mismatches 755; Indels 504; Gaps 83;

```
Qy      9 QRSPEQPAGSD----------TPSELVLSGHGPAAEASG--------AAGDPADADPAT- 49
Db    301 QVSQTEPAKSDLSKLESVRMKVPKEKGLSSHVEVVEKEGRLKARKHLKPEQPADGVSAVD 360

Qy     50 --KLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMA 107
Db    361 LEKLEARKRRFADSNLKAEKQKPEVKKSSPEMEDARVL-----SKKQPDVSSREVI---- 411

Qy    108 ANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTV 167
Db    412 --------LLREGEAERKPVRKEILKRES-----------------KKI---------- 435

Qy    168 KKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQE 227
Db    436 ------KLDRLNTVASPKDCQELASISVGSGSRPSSDLQARLGELAG------ESVENQE 483

Qy    228 GTLKQEQSSTEIP-------LQAESDQAAEEE--AKDEGEEKQEKEPTKSPESPSSPVNS 278
Db    484 -----VQSKKPIPSKPQLKQLQVLDDQGPEREDVRKNYCSLRDETPERKSGQEKSHSVNT 538

Qy    279 ET--------TSSFKKFFTHGWAGWRKKTSFKKSK-----EDDLETAEKR---------- 315
Db    539 EEKIGIDIDHTQSYRKQMEQSRRKQQMEMEIAKSEKFGSPKKDVDEYERRSLVHEVGKPP 598

Qy    316 ----------KEQEAEKVD-------EEEKEKTEPASEEQEPAEDTDQARLSADYEKVEL 358
Db    599 QDVTDDSPPSKKKRMDHVDFDICTKRERNYRSSRQISEDSERTGGSPSVRHGSFHED--- 655

Qy    359 PLEDQVGDLEASS-------EEKCAP----------LATEVFDEKMEAHQEVVAEVHVST 401
Db    656 --EDPIGSPRLLSVKGSPKVDEKVLPYSNITVREESLKFNPYDSSRREQMADMAKIKLSV 713

Qy    402 VEKTEEEQGGGGEAE---GGVVVEGTGESLPPEKL-------AEPQEVPQEAEPAEE--- 448
Db    714 LNSEDELNRWDSQMKQDAGRFDVSFPNSIIKRDSLRKRSVRDLEPGEVPSDSDEDGEHKS 773

Qy    449 ----------------LMKSREMCVSGGDHTQLTDLSPE-------EKTLPKHPEGIVSE 485
Db    774 HSPRASALYESSRLSFLLRDREDKLRERDERLSSSLERNKFYSFALDKTITPDTKALLER 833

Qy    486 VEMLSS---------------------QERIKVQGSP--------LKKLFSSSGLKKLSG 516
Db    834 AKSLSSSREENWSFLDWDSRFANFRNNKDKEKVDSAPRPIPSWYMKKKKIRTDSEGKMDD 893

Qy    517 KKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQD 576
Db    894 KKEDHK-----EEEQERQELFASRFLHSSIFEQDSKRLQHLERKEEDSDFISGRIYGKQT 948

Qy    577 GEAEEGTTSDGEKKREGITPWAS-FKKMV----TPKKRVRRPSESDKEEELEKVKSATLS 631
Db    949 SEGANSTT---DSIQEPVVLFHSRFMELTRMQQKEKEKDQKPKEVEKQEDTENHPKTPES 1005

Qy    632 STDSTVSEMQ-------DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASS 684
Db   1006 APENKDSELKTPPSVGPPSVTVVTLESAPSALEKTTGDKTVEAPLV---TEEKTVEPATV 1062

Qy    685 SDD-----EGGPRTLGGDSHRAEE------ASKDKEAGTDAVPASTQEQDQAQGSSSPEP 733
Db   1063 SEEAKPASEPAPAPV----EQLEQVDLPPGADPDKEAA--MMPAGVEE-----GSSGDQP 1111
```

```
Qy    734 ----AGSPSEGEGVSTWE--------SFKRLVTPRKKSKSKLEEKAE--DSSVE------ 773
Db   1112 PYLDAKPPTPGASFSQAESNVDPEPDSTQPLSKPAQKSEEANEPKAEKPDATADAEPDAN 1171

Qy    774 ---QLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVV 830
Db   1172 QKAEAAPESQPPASED-LEVDPPVAAKDKKPNKSKRSKTPVQ-AAAVSIVE-----KPVT 1224

Qy    831 PLSEYNAVEREKMEAQGNT--ELPQLLGAVYVSEELSKT--------LVHTVSVAVIDGT 880
Db   1225 RKSE--RIDREKLKRSNSPRGEAQKLLELKMEAEKITRTASKNSAADLEHPEPSLPLSRT 1282

Qy    881 RA--VTSV-------EERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAE--ETPVL 929
Db   1283 RRRNVRSVYATMGDHENRSP------VKEPVEQ---------PRVTRKRLERELQEAAAV 1327

Qy    930 TQT-----LPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAE----ETTD 980
Db   1328 PTTPRRGRPPKTRRRADEEEENEAKEPAETLKPPEGWRSPRSQKTAAGGGPQGKKGKNEP 1387

Qy    981 MVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQT 1040
Db   1388 KVDAT-----RPEATTEVGPQIGVKESSMEPKAAEEEAGSEQKRDRKDAGTDKNPPETAP 1442

Qy   1041 VQRTGSKALEKVEEVEEDS--------------EVLASEKEKD--VMPKGPVQEAGAEHL 1084
Db   1443 VE-----VVEKKPAPEKNSKSKRGRSRNSRLAVDKSASLKNVDAAVSPRGAAAQAGER-- 1495

Qy   1085 AQGSETG--QATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGS 1142
Db   1496 ----ESGVVAVSPEKSESPQKEDGLS----------SQLKSDPVDPDKEPEKEDVSASGP 1541

Qy   1143 TPLADSDTADGTQQDETIDS----QDSKATAAVRQS---------------QVTEEEAAT 1183
Db   1542 SPEA-TQLAKQMELEQAVEHIAKLAEASASAAYKADAPEGLAPEDRDKPAHQASETELAA 1600

Qy   1184 AQK---EEPSTLPNNVPAQEEH-GEE-------PGRDVLEPTQQEL-TAAAVPVLAKTEV 1231
Db   1601 AIGSIINDISGEPENFPAPPPYPGESQTDLQPPAGAQALQPSEEGMETDEAVSGILETEA 1660

Qy   1232 GQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLE 1291
Db   1661 ATESSRPPVNAPDPS--------AGPTDTKEARGNSSETSHSVPEAKGSKEVEVTLVRKD 1712

Qy   1292 EGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGG 1351
Db   1713 KGRQKTTRSRRKRNTNKKVVA---PVESHVPESNQAQG------ESPAANEGTTV----Q 1759

Qy   1352 SPSLPDQDKAGCIEVQVQSLDTTVTQTA-------EAVEKVIETVVISETGESPECVGAH 1404
Db   1760 HPEAPQEEK------QSEKPHSTPPQSCTSDLSKIPSTENSSQEISVEERTPTKASVPPD 1813

Qy   1405 L----LPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPA--------------- 1445
Db   1814 LPPPPQPAPVDEEPQARFRVHSIIESDPVTPPSD-PSIPIPTLPSVTAAKLSPPVASGGI 1872

Qy   1446 PESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKE----STAIEKVLKAEPEILELES 1501
Db   1873 PHQSPPTKVTEWITRQEEPRAQSTPSPALPPDTKASDVDTSSSTLRKIL-MDPKYVSATS 1931

Qy   1502 KSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHP 1561
Db   1932 VTSTSVTTAIAEPV------SAAPCLH------EAPPPPVDSKKP-------LEEKTAPP 1972

Qy   1562 VPQPREDLQVLTVLEAWAQPRKCLPRLQLK-------APVS 1595
Db   1973 VTN-NSEIQASEVLVA-ADKEKVAPVIAPKITSVISRMPVS 2011
```

RESULT 7
US-09-724-676-73643
; Sequence 73643, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73643
; LENGTH: 4368
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-73643

```
  Query Match             5.0%;  Score 401.5;  DB 5;  Length 4368;
  Best Local Similarity   19.9%;  Pred. No. 1.2e-14;
  Matches  317;  Conservative  272;  Mismatches  662;  Indels  339;  Gaps   63;

Qy      8 EQRSPEQPAGSD----TPSELVLSGHGPAAEASGAAGDPADADPATKLPQK------NGQ 57
Db   1374 EERLAEQQRAEERERLAEVEAALEKQRQLAEAHAQKAQAERE-AKELQQRMQEEVVRRE 1432

Qy     58 LSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESE-------DVREKDRVEEMAANS 110
Db   1433 EAAVD--AQQQKRSIQEELQQLRQSSEAEIQAKARQAEAAERSRLRIEEEIRVVRLQLEA 1490

Qy    111 TAVEDITKDGQEETSEI-IEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKK 169
Db   1491 TERQRGGAEGELQALRARAEEAEAQKRQAQEEAERLRRQVQDESQRKRQAEVELASRVKA 1550

Qy    170 D---KNEKSDTVQLLTVKKDEGEGAE-----ASVGAGDHQEPSVET----AVGESASKES 217
Db   1551 ETEAAREKQRALQALEELRLQAEEAERRLRQAEVERARQVQVALETAQRSAEAELQSKRA 1610

Qy    218 ELKQSTEKQEGTLKQEQ-SSTEIPLQAE---SDQAAEEEAKDEGEEKQEKEPTKSPESPS 273
Db   1611 SFAEKTAQLERSLQEEHVAVAQLREEAERRAQQQAEAERAREEAERELERWQLKANEALR 1670

Qy    274 SPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEP 333
Db   1671 LRLQAEEVAQQKSL---------AQAEAEKQKEEAEREARRRGKAEEQAVRQRELAEQE- 1720

Qy    334 ASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEAS-SEEKCAPL------ATEVFDE- 385
Db   1721 LEKQRQLAEGTAQQRLAAEQELIRLRAETEQGEQQRQLLEEELARLQREAAAATQKRQEL 1780

Qy    386 -----KMEAHQEVV-------AEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGE-----SL 428
Db   1781 EAELAKVRAEMEVLLASKARAEEESRSTSEKSKQRL----EAEAGRFRELAEEAARLRAL 1836

Qy    429 PPE-----KLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIV 483
Db   1837 AEEAKRQRQLAEEDAARQRAE-AERVLAEKLAAI--GEATRL-------KT-----EAEI 1881

Qy    484 SEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPE 543
Db   1882 ALKEKEAENER-------LRRLAEDEAFQRRRLEEQAAQHKADIEERLAQ---LRKASDS 1931

Qy    544 SADEQKG--ESSASSPEEPEETTCLEKGPLEAPQDGEAE--------EGTTSDGEKKREG 593
Db   1932 ELERQKGLVEDTLRQRRQVEEEILALKASFEKAAAGKAELELELGRIRSNAEDTLRSKEQ 1991

Qy    594 ITPWASFKKMVTPKKRVRRPSESDK-------EEELEKVKSATLSSTDSTVSEMQDEVKT 646
Db   1992 AELEAARQRQLAAEEERRRREAEERVQKSLAAEEEAARQRKAALEEVERLKANV-EEARR 2050

Qy    647 VGEEQKPEEPK----------RRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGG 696
Db   2051 LRERAEQESARQLQLAQEAAQKRLQAEEKAHAFAV--QQKEQELQQTLQQEQSVLDQLRG 2108

Qy    697 DSHRAEEASKDKE-----AGTDAVPASTQ-EQDQAQGSSSPEPAGSPSEGEGVSTWESFK 750
Db   2109 EAEAARRAAEEAEEARVQAEREAAQARRQVEEAERLKQSAEEQAQARAQAQAA------ 2162
```

```
Qy     751 RLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQAT 810
Db    2163 -----EKLRKEAEQEAARRAQAEQAALRQKQAADAEMEKHKKFAEQTLRQKAQVEQELTT 2217

Qy     811 VEDSGPVEINEDDPNVPAVVPLSEYNAVERE----KME----AQGNTELPQLLGAVYVS- 861
Db    2218 LR----LQLEETD---------HQKNLLDEELQRLKAEATEAARQRSQVEEELFSVRVQM 2264

Qy     862 EELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPV-EEVTEKD 920
Db    2265 EELSK-------------LKARIEAENRALILRDKDNTQRFLQEEAEKMKQVAEEAARLS 2311

Qy     921 IIAEETPVLTQ----TLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAE 976
Db    2312 VAAQEAARLRQLAEEDLAQQRALAEKMLKEKMQAVQEATRLKAEAELLQQQKELAQEQAR 2371

Qy     977 ETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK------ 1030
Db    2372 RLQE------------DKEQMAQQLAEETQGFQRTLEAERQRQLEMSAEAERLKLRVAEM 2419

Qy    1031 EESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSET 1090
Db    2420 SRAQARAEEDAQRFRKQAEEIGEKLHRTE--LATQEKVTLVQTLEIQRQQSDHDAE---- 2473

Qy    1091 GQATPESLEVPEVTADVDH--VATCQVIKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db    2474 --------RLREAIAELEREKEKLQQEAKLLQLKSEEMQTVQQEQLLQETQALQQSFLSE 2525

Qy    1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db    2526 KDSL--LQRERFIEQEKAKL------EQLFQDEVAKAQQ-----------LREE--QQRQ 2564

Qy    1208 RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTY 1267
Db    2565 QQQMEQERQRLVASMEEARRRQHEAEEG---------VRRKQE----------------- 2598

Qy    1268 DSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPE-QVSEEGEQETAAPEHEG 1326
Db    2599 ----------------ELQQLEQQRRQQEELLAEENQRLREQLQLLEEQHRAALAHSEEV 2642

Qy    1327 TYGKPVLTLDMPS-----------SERGKALGSLGGSPSLPDQDKAGCIEV-QVQSL--- 1371
Db    2643 TASQVAATKTLPNGRDALDGPAAEAEPEHSFDGLRRKVSAQRLQEAGILSAEELQRLAQG 2702

Qy    1372 DTTVTQTAEAVE--------KVIETVVISETGE-------------SPECVGAHLLPAEK 1410
Db    2703 HTTVDELARREDVRHYLQGRSSIAGLLLKATNEKLSVYAALQRQLLSP---GTALILLEA 2759

Qy    1411 SSATG--------GHWTLQHAEDTVPLGPE 1432
Db    2760 QAASGFLLDPVRNRRLTVNEAVKEGVVGPE 2789
```

RESULT 8
US-09-724-676A-73643
; Sequence 73643, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 73643
; LENGTH: 4368
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-73643

Query Match 5.0%; Score 401.5; DB 5; Length 4368;
Best Local Similarity 19.9%; Pred. No. 1.2e-14;
Matches 317; Conservative 272; Mismatches 662; Indels 339; Gaps 63;

```
Qy       8 EQRSPEQPAGSD----TPSELVLSGHGPAAEASGAAGDPADADPATKLPQK------NGQ 57
Db    1374 EERLAEQQRAEERERLAEVEAALEKQRQLAEAHAQAKAQAERE-AKELQQRMQEEVVRRE 1432

Qy      58 LSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESE-------DVREKDRVEEMAANS 110
Db    1433 EAAVD--AQQQKRSIQEELQQLRQSSEAEIQAKARQAEAAERSRLRIEEEIRVVRLQLEA 1490

Qy     111 TAVEDITKDGQEETSEI-IEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKK 169
Db    1491 TERQRGGAEGELQALRARAEEAEAQKRQAQEEAERLRRQVQDESQRKRQAEVELASRVKA 1550

Qy     170 D---KNEKSDTVQLLTVKKDEGEGAE-----ASVGAGDHQEPSVET----AVGESASKES 217
Db    1551 ETEAAREKQRALQALEELRLQAEEAERRLRQAEVERARQVQVALETAQRSAEAELQSKRA 1610

Qy     218 ELKQSTEKQEGTLKQEQ-SSTEIPLQAE---SDQAAEEEAKDEGEEKQEKEPTKSPESPS 273
Db    1611 SFAEKTAQLERSLQEEHVAVAQLREEAERRAQQQAEAERAREEAERELERWQLKANEALR 1670

Qy     274 SPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEP 333
Db    1671 LRLQAEEVAQQKSL---------AQAEAEKQKEEAEREARRRGKAEEQAVRQRELAEQE- 1720

Qy     334 ASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEAS-SEEKCAPL------ATEVFDE- 385
Db    1721 LEKQRQLAEGTAQQRLAAEQELIRLRAETEQGEQQRQLLEEELARLQREAAAATQKRQEL 1780

Qy     386 -----KMEAHQEVV-------AEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGE-----SL 428
Db    1781 EAELAKVRAEMEVLLASKARAEEESRSTSEKSKQRL----EAEAGRFRELAEEAARLRAL 1836

Qy     429 PPE-----KLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIV 483
Db    1837 AEEAKRQRQLAEEDAARQRAE-AERVLAEKLAAI--GEATRL-------KT-----EAEI 1881

Qy     484 SEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPE 543
Db    1882 ALKEKEAENER-------LRRLAEDEAFQRRRLEEQAAQHKADIEERLAQ---LRKASDS 1931

Qy     544 SADEQKG--ESSASSPEEPEETTCLEKGPLEAPQDGEAE--------EGTTSDGEKKREG 593
Db    1932 ELERQKGLVEDTLRQRRQVEEEILALKASFEKAAAGKAELELELGRIRSNAEDTLRSKEQ 1991

Qy     594 ITPWASFKKMVTPKKRVRRPSESDK-------EEELEKVKSATLSSTDSTVSEMQDEVKT 646
Db    1992 AELEAARQRQLAAEEERRRREAEERVQKSLAAEEEAARQRKAALEEVERLKANV-EEARR 2050

Qy     647 VGEEQKPEEPK----------RRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGG 696
Db    2051 LRERAEQESARQLQLAQEAAQKRLQAEEKAHAFAV--QQKEQELQQTLQQEQSVLDQLRG 2108

Qy     697 DSHRAEEASKDKE-----AGTDAVPASTQ-EQDQAQGSSSPEPAGSPSEGEGVSTWESFK 750
Db    2109 EAEAARRAAEEAEEARVQAEREAAQARRQVEEAERLKQSAEEQAQARAQAQAAA------ 2162

Qy     751 RLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQAT 810
Db    2163 -----EKLRKEAEQEAARRAQAEQAALRQKQAADAEMEKHKKFAEQTLRQKAQVEQELTT 2217

Qy     811 VEDSGPVEINEDDPNVPAVVPLSEYNAVERE----KME----AQGNTELPQLLGAVYVS- 861
Db    2218 LR----LQLEETD---------HQKNLLDEELQRLKAEATEAARQRSQVEEELFSVRVQM 2264

Qy     862 EELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPV-EEVTEKD 920
Db    2265 EELSK-------------LKARIEAENRALILRDKDNTQRFLQEEAEKMKQVAEEAARLS 2311

Qy     921 IIAEETPVLTQ----TLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAE 976
Db    2312 VAAQEAARLRQLAEEDLAQQRALAEKMLKEKMQAVQEATRLKAEAELLQQQKELAQEQAR 2371
```

```
Qy    977 ETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK------ 1030
Db   2372 RLQE-----------DKEQMAQQLAEETQGFQRTLEAERQRQLEMSAEAERLKLRVAEM 2419

Qy   1031 EESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSET 1090
Db   2420 SRAQARAEEDAQRFRKQAEEIGEKLHRTE--LATQEKVTLVQTLEIQRQQSDHDAE---- 2473

Qy   1091 GQATPESLEVPEVTADVDH--VATCQVIKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db   2474 --------RLREAIAELEREKEKLQQEAKLLQLKSEEMQTVQQEQLLQETQALQQSFLSE 2525

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db   2526 KDSL--LQRERFIEQEKAKL------EQLFQDEVAKAQQ-----------LREE--QQRQ 2564

Qy   1208 RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTY 1267
Db   2565 QQQMEQERQRLVASMEEARRRQHEAEEG---------VRRKQE----------------- 2598

Qy   1268 DSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPE-QVSEEGEQETAAPEHEG 1326
Db   2599 ----------------ELQQLEQQRRQQEELLAEENQRLREQLQLLEEQHRAALAHSEEV 2642

Qy   1327 TYGKPVLTLDMPS-----------SERGKALGSLGGSPSLPDQDKAGCIEV-QVQSL--- 1371
Db   2643 TASQVAATKTLPNGRDALDGPAAEAEPEHSFDGLRRKVSAQRLQEAGILSAEELQRLAQG 2702

Qy   1372 DTTVTQTAEAVE--------KVIETVVISETGE-------------SPECVGAHLLPAEK 1410
Db   2703 HTTVDELARREDVRHYLQGRSSIAGLLLKATNEKLSVYAALQRQLLSP---GTALILLEA 2759

Qy   1411 SSATG--------GHWTLQHAEDTVPLGPE 1432
Db   2760 QAASGFLLDPVRNRRLTVNEAVKEGVVGPE 2789
```

RESULT 9
US-09-724-676-55851
; Sequence 55851, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 55851
;   LENGTH: 2642
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676-55851

```
  Query Match             5.0%;  Score 400;  DB 5;  Length 2642;
  Best Local Similarity   19.1%;  Pred. No. 6.8e-15;
  Matches  306;  Conservative  304;  Mismatches  663;  Indels  330;  Gaps   58;

Qy     43 ADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR 102
Db    738 AEAERQRRLDYES-QTAHDNLLTEQ--IHSLSIEAKSKDVKI---EVLQNELDDVQ---- 787

Qy    103 VEEMAANSTAVEDITKDGQEETSEII---EQIPASENNVEEMVQPAESQANDVGFKKVFK 159
Db    788 -LQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEELSQALSQKELEI------- 839

Qy    160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESEL 219
Db    840 -------TKMD--------QLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEE 884

Qy    220 KQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSE 279
Db    885 KFSLGVEIKTLKE-----QLNLLSRAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKE 939

Qy    280 TTSSFKKFFTHGWAGWRKKTSFKKSK--------EDDLETAEKRKEQEAEKVDEEEKEKT 331
Db    940 E-------LQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIP 992

Qy    332 EPASEEQEPAEDTDQARLS-----ADYEKVELPLEDQVGDLEA-------SSEEKCAPLA 379
Db    993 LSETERGEVEEDKENKEYSEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEK---LA 1049

Qy    380 TEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEV 439
Db   1050 AE------EQFQALVKQMNQTLQDKTNQ------------------IDLLQAEISENQAI 1085

Qy    440 PQEAEPAEELMKSREMCVSGGDHTQLTD----LSPEEKTLPKHPEGIVSEVEMLSSQERI 495
Db   1086 IQK-------LITSNTDASDGDSVALVKETVVISPPCTGSSEHWKPELEEKILALEKEKE 1138

Qy    496 KVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGES--- 552
Db   1139 QLQKKLQEALTSRKAILKKAQEKERHLR-----EELKQQKDDYNRLQEQFDEQSKENENI 1193

Qy    553 ---------------SASSPEEPEETTCLEKGPLEAPQDGEAEEGTT------------- 584
Db   1194 GDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDWP 1253

Qy    585 --SDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK-SATLSSTDSTVSEMQ 641
Db   1254 SHSEDASALQGGTSVAQIKAQL-------------KEIEAEKVELELKVSSTTSELTKKS 1300

Qy    642 DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRA 701
Db   1301 EEVFQLQEQINKQGLEIESLKTVSHEAEVHAESLQQKL---------ESSQLQIAGLEHLR 1352

Qy    702 EEASKDKEAGTDAVPASTQEQDQAQGS---SSPEPAGSPSEGEGVSTWESFKRLVTPRKK 758
Db   1353 ELQPKLDEL---QKLISKKEEDVSYLSGQLSEKEAALTKIQTEIIEQEDLIKALHT---- 1405

Qy    759 SKSKLEEKAEDSSVEQLSTEI-------EPSREESWVSIKKFIPGRRKKRADGKQEQATV 811
Db   1406 -QLEMQAKEHDERIKQLQVELCEMKQKPEEIGEES----------RAKQQIQRKLQAALI 1454

Qy    812 EDSGPVEINEDDPNVPAVVPLSEYNAVER--------EKMEAQGNTELPQLLGAVYVSEE 863
Db   1455 SRK---EALKENKSLQEELSLAR-GTIERLTKSLADVESQVSAQNKEKDTVLGRLALLQE 1510

Qy    864 LSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA 923
Db   1511 ERDKLITEMDRSLLENQSLSSSCE--SLKLALEGLTEDKEKLVKEI-----ESLKSSKIA 1563

Qy    924 EETPVLTQTLPEGKDAHDD-------MVTSEVDFTSEAVTATETSEALRTEE-------- 968
Db   1564 EST--------EWQEKHKELQKEYEILLQSYENVSNEAERIQHVVEAVRQEKQELYGKLR 1615

Qy    969 VTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADK 1028
Db   1616 STEANKKETEKQLQEAEQEMEEMKEKMRKF--AKSKQQKILELEEENDRLRAEVHPAGDT 1673

Qy   1029 VKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASE-----KEKDVMPKGPVQEAGAEH 1083
Db   1674 AKE-----CMETLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSE-EVQD--LKH 1725

Qy   1084 LAQGSETGQATPESLEVPEVTADVDHVATCQV---IKLQQLMEQAVAPESSETLTDSETN 1140
Db   1726 QIEGNVSKQANLEATEKHDNQTNVTEEGTQSIPGETEEQDSLSMSTRPTCSESVPSAKS- 1784

Qy   1141 GSTPLADSDTADGTQQD---ETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVP 1197
Db   1785 -ANPAVSKDFSSHDEINNYLQQIDQLKERIAGLEEEKQKNKEFSQTLENEK-NTLLSQIS 1842

Qy   1198 AQEEHGEEPGRDVLEPTQQELT---------AAAVPVLAKTEVGQEGEVDWLDGEKVKEE 1248
Db   1843 TKD--GE------LKMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKDDLEERLMNQL 1894
```

```
Qy   1249 QEVFVHSGPNSQKAADVT--------YDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVE 1300
Db   1895 AEL---NGSIGNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVKEKTKVESEIR 1951

Qy   1301 KEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDK 1360
Db   1952 KEYLEKIQGAQKEPGNK---------SHAKELQELLKEKQQEVKQL----QKDCIRYQEK 1998

Qy   1361 AGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH------LLPAEKSSAT 1414
Db   1999 ISALERTVKALEFVQTESQKDLEITKEN--LAQAVEHRKKAQAELASFKVLLDDTQSEAA 2056

Qy   1415 GGHWTLQHAEDTVPLGPESQA--ESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKP 1472
Db   2057 ------RVLADNLKLKKELQSNKESVKSQMKQKDE-----DLERRLEQAEEKHLKEKKNM 2105

Qy   1473 DAGPDADGKESTAIEKVL--------KAEPEILELESKSNKIV 1507
Db   2106 QEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQENLDSTV 2148
```

RESULT 10
US-09-724-676A-55851
; Sequence 55851, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 55851
; LENGTH: 2642
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676A-55851

```
  Query Match             5.0%;  Score 400;  DB 5;  Length 2642;
  Best Local Similarity   19.1%;  Pred. No. 6.8e-15;
  Matches  306;  Conservative  304;  Mismatches  663;  Indels  330;  Gaps   58;

Qy     43 ADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR 102
Db    738 AEAERQRRLDYES-QTAHDNLLTEQ--IHSLSIEAKSKDVKI---EVLQNELDDVQ---- 787

Qy    103 VEEMAANSTAVEDITKDGQEETSEII---EQIPASENNVEEMVQPAESQANDVGFKKVFK 159
Db    788 -LQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEELSQALSQKELEI------- 839

Qy    160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESEL 219
Db    840 -------TKMD--------QLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEE 884

Qy    220 KQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSE 279
Db    885 KFSLGVEIKTLKE-----QLNLLSRAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKE 939

Qy    280 TTSSFKKFFTHGWAGWRKKTSFKKSK--------EDDLETAEKRKEQEAEKVDEEEKEKT 331
Db    940 E-------LQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIP 992

Qy    332 EPASEEQEPAEDTDQARLS-----ADYEKVELPLEDQVGDLEA-------SSEEKCAPLA 379
Db    993 LSETERGEVEEDKENKEYSEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEK---LA 1049

Qy    380 TEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEV 439
Db   1050 AE------EQFQALVKQMNQTLQDKTNQ------------------IDLLQAEISENQAI 1085

Qy    440 PQEAEPAEELMKSREMCVSGGDHTQLTD----LSPEEKTLPKHPEGIVSEVEMLSSQERI 495
Db   1086 IQK-------LITSNTDASDGDSVALVKETVVISPPCTGSSEHWKPELEEKILALEKEKE 1138

Qy    496 KVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGES--- 552
Db   1139 QLQKKLQEALTSRKAILKKAQEKERHLR-----EELKQQKDDYNRLQEQFDEQSKENENI 1193

Qy    553 ---------------SASSPEEPEETTCLEKGPLEAPQDGEAEEGTT------------- 584
Db   1194 GDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDWP 1253

Qy    585 --SDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK-SATLSSTDSTVSEMQ 641
Db   1254 SHSEDASALQGGTSVAQIKAQL-------------KEIEAEKVELELKVSSTTSELTKKS 1300

Qy    642 DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRA 701
Db   1301 EEVFQLQEQINKQGLEIESLKTVSHEAEVHAESLQQKL--------ESSQLQIAGLEHLR 1352

Qy    702 EEASKDKEAGTDAVPASTQEQDQAQGS---SSPEPAGSPSEGEGVSTWESFKRLVTPRKK 758
Db   1353 ELQPKLDEL---QKLISKKEEDVSYLSGQLSEKEAALTKIQTEIIEQEDLIKALHT---- 1405

Qy    759 SKSKLEEKAEDSSVEQLSTEI-------EPSREESWVSIKKFIPGRRKKRADGKQEQATV 811
Db   1406 -QLEMQAKEHDERIKQLQVELCEMKQKPEEIGEES----------RAKQQIQRKLQAALI 1454

Qy    812 EDSGPVEINEDDPNVPAVVPLSEYNAVER--------EKMEAQGNTELPQLLGAVYVSEE 863
Db   1455 SRK---EALKENKSLQEELSLAR-GTIERLTKSLADVESQVSAQNKEKDTVLGRLALLQE 1510

Qy    864 LSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA 923
Db   1511 ERDKLITEMDRSLLENQSLSSSCE--SLKLALEGLTEDKEKLVKEI-----ESLKSSKIA 1563

Qy    924 EETPVLTQTLPEGKDAHDD-------MVTSEVDFTSEAVTATETSEALRTEE-------- 968
Db   1564 EST--------EWQEKHKELQKEYEILLQSYENVSNEAERIQHVVEAVRQEKQELYGKLR 1615

Qy    969 VTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADK 1028
Db   1616 STEANKKETEKQLQEAEQEMEEMKEKMRKF--AKSKQQKILELEEENDRLRAEVHPAGDT 1673

Qy   1029 VKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASE-----KEKDVMPKGPVQEAGAEH 1083
Db   1674 AKE-----CMETLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSE-EVQD--LKH 1725

Qy   1084 LAQGSETGQATPESLEVPEVTADVDHVATCQV---IKLQQLMEQAVAPESSETLTDSETN 1140
Db   1726 QIEGNVSKQANLEATEKHDNQTNVTEEGTQSIPGETEEQDSLSMSTRPTCSESVPSAKS- 1784

Qy   1141 GSTPLADSDTADGTQQD---ETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVP 1197
Db   1785 -ANPAVSKDFSSHDEINNYLQQIDQLKERIAGLEEEKQKNKEFSQTLENEK-NTLLSQIS 1842

Qy   1198 AQEEHGEEPGRDVLEPTQQELT---------AAAVPVLAKTEVGQEGEVDWLDGEKVKEE 1248
Db   1843 TKD--GE------LKMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKDDLEERLMNQL 1894

Qy   1249 QEVFVHSGPNSQKAADVT--------YDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVE 1300
Db   1895 AEL---NGSIGNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVKEKTKVESEIR 1951

Qy   1301 KEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDK 1360
Db   1952 KEYLEKIQGAQKEPGNK---------SHAKELQELLKEKQQEVKQL----QKDCIRYQEK 1998

Qy   1361 AGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH------LLPAEKSSAT 1414
Db   1999 ISALERTVKALEFVQTESQKDLEITKEN--LAQAVEHRKKAQAELASFKVLLDDTQSEAA 2056

Qy   1415 GGHWTLQHAEDTVPLGPESQA--ESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKP 1472
Db   2057 ------RVLADNLKLKKELQSNKESVKSQMKQKDE-----DLERRLEQAEEKHLKEKKNM 2105
```

```
Qy   1473 DAGPDADGKESTAIEKVL--------KAEPEILELESKSNKIV 1507
Db   2106 QEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQENLDSTV 2148
```

RESULT 11
US-09-724-676-55868
; Sequence 55868, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 55868
;   LENGTH: 2722
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676-55868

```
  Query Match             5.0%;  Score 400;  DB 5;  Length 2722;
  Best Local Similarity   19.1%;  Pred. No. 7.1e-15;
  Matches  306;  Conservative  304;  Mismatches  663;  Indels  330;  Gaps   58;

Qy     43 ADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR 102
Db    818 AEAERQRRLDYES-QTAHDNLLTEQ--IHSLSIEAKSKDVKI---EVLQNELDDVQ---- 867

Qy    103 VEEMAANSTAVEDITKDGQEETSEII---EQIPASENNVEEMVQPAESQANDVGFKKVFK 159
Db    868 -LQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEELSQALSQKELEI------- 919

Qy    160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESEL 219
Db    920 -------TKMD--------QLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEE 964

Qy    220 KQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSE 279
Db    965 KFSLGVEIKTLKE-----QLNLLSRAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKE 1019

Qy    280 TTSSFKKFFTHGWAGWRKKTSFKKSK--------EDDLETAEKRKEQEAEKVDEEEKEKT 331
Db   1020 E-------LQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIP 1072

Qy    332 EPASEEQEPAEDTDQARLS-----ADYEKVELPLEDQVGDLEA-------SSEEKCAPLA 379
Db   1073 LSETERGEVEEDKENKEYSEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEK---LA 1129

Qy    380 TEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEV 439
Db   1130 AE------EQFQALVKQMNQTLQDKTNQ-----------------IDLLQAEISENQAI 1165

Qy    440 PQEAEPAEELMKSREMCVSGGDHTQLTD----LSPEEKTLPKHPEGIVSEVEMLSSQERI 495
Db   1166 IQK-------LITSNTDASDGDSVALVKETVVISPPCTGSSEHWKPELEEKILALEKEKE 1218

Qy    496 KVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGES--- 552
Db   1219 QLQKKLQEALTSRKAILKKAQEKERHLR-----EELKQQKDDYNRLQEQFDEQSKENENI 1273

Qy    553 ---------------SASSPEEPEETTCLEKGPLEAPQDGEAEEGTT------------- 584
Db   1274 GDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDWP 1333

Qy    585 --SDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK-SATLSSTDSTVSEMQ 641
Db   1334 SHSEDASALQGGTSVAQIKAQL-------------KEIEAEKVELELKVSSTTSELTKKS 1380

Qy    642 DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRA 701
Db   1381 EEVFQLQEQINKQGLEIESLKTVSHEAEVHAESLQQKL--------ESSQLQIAGLEHLR 1432

Qy    702 EEASKDKEAGTDAVPASTQEQDQAQGS---SSPEPAGSPSEGEGVSTWESFKRLVTPRKK 758
Db   1433 ELQPKLDEL---QKLISKKEEDVSYLSGQLSEKEAALTKIQTEIIEQEDLIKALHT---- 1485

Qy    759 SKSKLEEKAEDSSVEQLSTEI-------EPSREESWVSIKKFIPGRRKKRADGKQEQATV 811
Db   1486 -QLEMQAKEHDERIKQLQVELCEMKQKPEEIGEES---------RAKQQIQRKLQAALI 1534

Qy    812 EDSGPVEINEDDPNVPAVVPLSEYNAVER--------EKMEAQGNTELPQLLGAVYVSEE 863
Db   1535 SRK---EALKENKSLQEELSLAR-GTIERLTKSLADVESQVSAQNKEKDTVLGRLALLQE 1590

Qy    864 LSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA 923
Db   1591 ERDKLITEMDRSLLENQSLSSSCE--SLKLALEGLTEDKEKLVKEI-----ESLKSSKIA 1643

Qy    924 EETPVLTQTLPEGKDAHDD-------MVTSEVDFTSEAVTATETSEALRTEE-------- 968
Db   1644 EST--------EWQEKHKELQKEYEILLQSYENVSNEAERIQHVVEAVRQEKQELYGKLR 1695

Qy    969 VTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADK 1028
Db   1696 STEANKKETEKQLQEAEQEMEEMKEKMRKF--AKSKQQKILELEEENDRLRAEVHPAGDT 1753

Qy   1029 VKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASE-----KEKDVMPKGPVQEAGAEH 1083
Db   1754 AKE-----CMETLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSE-EVQD--LKH 1805

Qy   1084 LAQGSETGQATPESLEVPEVTADVDHVATCQV---IKLQQLMEQAVAPESSETLTDSETN 1140
Db   1806 QIEGNVSKQANLEATEKHDNQTNVTEEGTQSIPGETEEQDSLSMSTRPTCSESVPSAKS- 1864

Qy   1141 GSTPLADSDTADGTQQD---ETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVP 1197
Db   1865 -ANPAVSKDFSSHDEINNYLQQIDQLKERIAGLEEEKQKNKEFSQTLENEK-NTLLSQIS 1922

Qy   1198 AQEEHGEEPGRDVLEPTQQELT---------AAAVPVLAKTEVGQEGEVDWLDGEKVKEE 1248
Db   1923 TKD--GE------LKMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKDDLEERLMNQL 1974

Qy   1249 QEVFVHSGPNSQKAADVT--------YDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVE 1300
Db   1975 AEL---NGSIGNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVKEKTKVESEIR 2031

Qy   1301 KEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDK 1360
Db   2032 KEYLEKIQGAQKEPGNK---------SHAKELQELLKEKQQEVKQL----QKDCIRYQEK 2078

Qy   1361 AGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH------LLPAEKSSAT 1414
Db   2079 ISALERTVKALEFVQTESQKDLEITKEN--LAQAVEHRKKAQAELASFKVLLDDTQSEAA 2136

Qy   1415 GGHWTLQHAEDTVPLGPESQA--ESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKP 1472
Db   2137 ------RVLADNLKLKKELQSNKESVKSQMKQKDE-----DLERRLEQAEEKHLKEKKNM 2185

Qy   1473 DAGPDADGKESTAIEKVL--------KAEPEILELESKSNKIV 1507
Db   2186 QEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQENLDSTV 2228
```

RESULT 12
US-09-724-676A-55868
; Sequence 55868, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2

```
; SEQ ID NO 55868
;   LENGTH: 2722
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676A-55868

  Query Match             5.0%;  Score 400;  DB 5;  Length 2722;
  Best Local Similarity   19.1%;  Pred. No. 7.1e-15;
  Matches  306;  Conservative  304;  Mismatches  663;  Indels  330;  Gaps   58;

Qy       43 ADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR 102
Db      818 AEAERQRRLDYES-QTAHDNLLTEQ--IHSLSIEAKSKDVKI---EVLQNELDDVQ---- 867

Qy      103 VEEMAANSTAVEDITKDGQEETSEII---EQIPASENNVEEMVQPAESQANDVGFKKVFK 159
Db      868 -LQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEELSQALSQKELEI------- 919

Qy      160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESEL 219
Db      920 -------TKMD--------QLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEE 964

Qy      220 KQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSE 279
Db      965 KFSLGVEIKTLKE-----QLNLLSRAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKE 1019

Qy      280 TTSSFKKFFTHGWAGWRKKTSFKKSK--------EDDLETAEKRKEQEAEKVDEEEKEKT 331
Db     1020 E-------LQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIP 1072

Qy      332 EPASEEQEPAEDTDQARLS-----ADYEKVELPLEDQVGDLEA-------SSEEKCAPLA 379
Db     1073 LSETERGEVEEDKENKEYSEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEK---LA 1129

Qy      380 TEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEV 439
Db     1130 AE------EQFQALVKQMNQTLQDKTNQ-----------------IDLLQAEISENQAI 1165

Qy      440 PQEAEPAEELMKSREMCVSGGDHTQLTD----LSPEEKTLPKHPEGIVSEVEMLSSQERI 495
Db     1166 IQK-------LITSNTDASDGDSVALVKETVVISPPCTGSSEHWKPELEEKILALEKEKE 1218

Qy      496 KVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGES--- 552
Db     1219 QLQKKLQEALTSRKAILKKAQEKERHLR-----EELKQQKDDYNRLQEQFDEQSKENENI 1273

Qy      553 ---------------SASSPEEPEETTCLEKGPLEAPQDGEAEEGTT------------- 584
Db     1274 GDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDWP 1333

Qy      585 --SDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK-SATLSSTDSTVSEMQ 641
Db     1334 SHSEDASALQGGTSVAQIKAQL-------------KEIEAEKVELELKVSSTTSELTKKS 1380

Qy      642 DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRA 701
Db     1381 EEVFQLQEQINKQGLEIESLKTVSHEAEVHAESLQQKL--------ESSQLQIAGLEHLR 1432

Qy      702 EEASKDKEAGTDAVPASTQEQDQAQGS---SSPEPAGSPSEGEGVSTWESFKRLVTPRKK 758
Db     1433 ELQPKLDEL---QKLISKKEEDVSYLSGQLSEKEAALTKIQTEIIEQEDLIKALHT---- 1485

Qy      759 SKSKLEEKAEDSSVEQLSTEI-------EPSREESWVSIKKFIPGRRKKRADGKQEQATV 811
Db     1486 -QLEMQAKEHDERIKQLQVELCEMKQKPEEIGEES----------RAKQQIQRKLQAALI 1534

Qy      812 EDSGPVEINEDDPNVPAVVPLSEYNAVER--------EKMEAQGNTELPQLLGAVYVSEE 863
Db     1535 SRK---EALKENKSLQEELSLAR-GTIERLTKSLADVESQVSAQNKEKDTVLGRLALLQE 1590

Qy      864 LSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA 923
Db     1591 ERDKLITEMDRSLLENQSLSSSCE--SLKLALEGLTEDKEKLVKEI-----ESLKSSKIA 1643

Qy      924 EETPVLTQTLPEGKDAHDD-------MVTSEVDFTSEAVTATETSEALRTEE-------- 968
Db     1644 EST--------EWQEKHKELQKEYEILLQSYENVSNEAERIQHVVEAVRQEKQELYGKLR 1695

Qy      969 VTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADK 1028
Db     1696 STEANKKETEKQLQEAEQEMEEMKEKMRKF--AKSKQQKILELEEENDRLRAEVHPAGDT 1753

Qy     1029 VKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASE-----KEKDVMPKGPVQEAGAEH 1083
Db     1754 AKE-----CMETLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSE-EVQD--LKH 1805

Qy     1084 LAQGSETGQATPESLEVPEVTADVDHVATCQV---IKLQQLMEQAVAPESSETLTDSETN 1140
Db     1806 QIEGNVSKQANLEATEKHDNQTNVTEEGTQSIPGETEEQDSLSMSTRPTCSESVPSAKS- 1864

Qy     1141 GSTPLADSDTADGTQQD---ETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVP 1197
Db     1865 -ANPAVSKDFSSHDEINNYLQQIDQLKERIAGLEEEKQKNKEFSQTLENEK-NTLLSQIS 1922

Qy     1198 AQEEHGEEPGRDVLEPTQQELT---------AAAVPVLAKTEVGQEGEVDWLDGEKVKEE 1248
Db     1923 TKD--GE------LKMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKDDLEERLMNQL 1974

Qy     1249 QEVFVHSGPNSQKAADVT--------YDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVE 1300
Db     1975 AEL---NGSIGNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVKEKTKVESEIR 2031

Qy     1301 KEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDK 1360
Db     2032 KEYLEKIQGAQKEPGNK---------SHAKELQELLKEKQQEVKQL----QKDCIRYQEK 2078

Qy     1361 AGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH------LLPAEKSSAT 1414
Db     2079 ISALERTVKALEFVQTESQKDLEITKEN--LAQAVEHRKKAQAELASFKVLLDDTQSEAA 2136

Qy     1415 GGHWTLQHAEDTVPLGPESQA--ESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKP 1472
Db     2137 ------RVLADNLKLKKELQSNKESVKSQMKQKDE-----DLERRLEQAEEKHLKEKKNM 2185

Qy     1473 DAGPDADGKESTAIEKVL--------KAEPEILELESKSNKIV 1507
Db     2186 QEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQENLDSTV 2228

RESULT 13
US-09-724-676-55849
; Sequence 55849, Application US/09724676
; GENERAL INFORMATION:
;  APPLICANT: Compugen LTD
;  TITLE OF INVENTION: Variants of alternative splicing
;  FILE REFERENCE: 129181.4 Compugen
;  CURRENT APPLICATION NUMBER: US/09/724,676
;  CURRENT FILING DATE:  2000-11-28
;  NUMBER OF SEQ ID NOS: 97222
;  SOFTWARE: PatentIn version 3.2
; SEQ ID NO 55849
;   LENGTH: 2915
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676-55849

  Query Match             5.0%;  Score 400;  DB 5;  Length 2915;
  Best Local Similarity   19.1%;  Pred. No. 7.9e-15;
  Matches  306;  Conservative  304;  Mismatches  663;  Indels  330;  Gaps   58;

Qy       43 ADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR 102
Db      738 AEAERQRRLDYES-QTAHDNLLTEQ--IHSLSIEAKSKDVKI---EVLQNELDDVQ---- 787

Qy      103 VEEMAANSTAVEDITKDGQEETSEII---EQIPASENNVEEMVQPAESQANDVGFKKVFK 159
```

```
Db    788 -LQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEELSQALSQKELEI------- 839

Qy    160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESEL 219
Db    840 -------TKMD--------QLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEE 884

Qy    220 KQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSE 279
Db    885 KFSLGVEIKTLKE-----QLNLLSRAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKE 939

Qy    280 TTSSFKKFFTHGWAGWRKKTSFKKSK--------EDDLETAEKRKEQEAEKVDEEEKEKT 331
Db    940 E-------LQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIP 992

Qy    332 EPASEEQEPAEDTDQARLS-----ADYEKVELPLEDQVGDLEA-------SSEEKCAPLA 379
Db    993 LSETERGEVEEDKENKEYSEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEK---LA 1049

Qy    380 TEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEV 439
Db   1050 AE------EQFQALVKQMNQTLQDKTNQ-----------------IDLLQAEISENQAI 1085

Qy    440 PQEAEPAEELMKSREMCVSGGDHTQLTD----LSPEEKTLPKHPEGIVSEVEMLSSQERI 495
Db   1086 IQK-------LITSNTDASDGDSVALVKETVVISPPCTGSSEHWKPELEEKILALEKEKE 1138

Qy    496 KVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGES--- 552
Db   1139 QLQKKLQEALTSRKAILKKAQEKERHLR-----EELKQQKDDYNRLQEQFDEQSKENENI 1193

Qy    553 ---------------SASSPEEPEETTCLEKGPLEAPQDGEAEEGTT------------- 584
Db   1194 GDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDWP 1253

Qy    585 --SDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK-SATLSSTDSTVSEMQ 641
Db   1254 SHSEDASALQGGTSVAQIKAQL-------------KEIEAEKVELELKVSSTTSELTKKS 1300

Qy    642 DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRA 701
Db   1301 EEVFQLQEQINKQGLEIESLKTVSHEAEVHAESLQQKL--------ESSQLQIAGLEHLR 1352

Qy    702 EEASKDKEAGTDAVPASTQEQDQAQGS---SSPEPAGSPSEGEGVSTWESFKRLVTPRKK 758
Db   1353 ELQPKLDEL---QKLISKKEEDVSYLSGQLSEKEAALTKIQTEIIEQEDLIKALHT---- 1405

Qy    759 SKSKLEEKAEDSSVEQLSTEI-------EPSREESWVSIKKFIPGRRKKRADGKQEQATV 811
Db   1406 -QLEMQAKEHDERIKQLQVELCEMKQKPEEIGEES----------RAKQQIQRKLQAALI 1454

Qy    812 EDSGPVEINEDDPNVPAVVPLSEYNAVER--------EKMEAQGNTELPQLLGAVYVSEE 863
Db   1455 SRK---EALKENKSLQEELSLAR-GTIERLTKSLADVESQVSAQNKEKDTVLGRLALLQE 1510

Qy    864 LSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA 923
Db   1511 ERDKLITEMDRSLLENQSLSSSCE--SLKLALEGLTEDKEKLVKEI-----ESLKSSKIA 1563

Qy    924 EETPVLTQTLPEGKDAHDD-------MVTSEVDFTSEAVTATETSEALRTEE-------- 968
Db   1564 EST--------EWQEKHKELQKEYEILLQSYENVSNEAERIQHVVEAVRQEKQELYGKLR 1615

Qy    969 VTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADK 1028
Db   1616 STEANKKETEKQLQEAEQEMEEMKEKMRKF--AKSKQQKILELEEENDRLRAEVHPAGDT 1673

Qy   1029 VKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASE-----KEKDVMPKGPVQEAGAEH 1083
Db   1674 AKE-----CMETLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSE-EVQD--LKH 1725

Qy   1084 LAQGSETGQATPESLEVPEVTADVDHVATCQV---IKLQQLMEQAVAPESSETLTDSETN 1140
Db   1726 QIEGNVSKQANLEATEKHDNQTNVTEEGTQSIPGETEEQDSLSMSTRPTCSESVPSAKS- 1784

Qy   1141 GSTPLADSDTADGTQQD---ETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVP 1197
Db   1785 -ANPAVSKDFSSHDEINNYLQQIDQLKERIAGLEEEKQKNKEFSQTLENEK-NTLLSQIS 1842

Qy   1198 AQEEHGEEPGRDVLEPTQQELT---------AAAVPVLAKTEVGQEGEVDWLDGEKVKEE 1248
Db   1843 TKD--GE------LKMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKDDLEERLMNQL 1894

Qy   1249 QEVFVHSGPNSQKAADVT--------YDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVE 1300
Db   1895 AEL---NGSIGNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVKEKTKVESEIR 1951

Qy   1301 KEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDK 1360
Db   1952 KEYLEKIQGAQKEPGNK---------SHAKELQELLKEKQQEVKQL----QKDCIRYQEK 1998

Qy   1361 AGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH------LLPAEKSSAT 1414
Db   1999 ISALERTVKALEFVQTESQKDLEITKEN--LAQAVEHRKKAQAELASFKVLLDDTQSEAA 2056

Qy   1415 GGHWTLQHAEDTVPLGPESQA--ESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKP 1472
Db   2057 ------RVLADNLKLKKELQSNKESVKSQMKQKDE-----DLERRLEQAEEKHLKEKKNM 2105

Qy   1473 DAGPDADGKESTAIEKVL--------KAEPEILELESKSNKIV 1507
Db   2106 QEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQENLDSTV 2148
```

RESULT 14
US-09-724-676-55850
; Sequence 55850, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 55850
; LENGTH: 2915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-724-676-55850

Query Match 5.0%; Score 400; DB 5; Length 2915;
Best Local Similarity 19.1%; Pred. No. 7.9e-15;
Matches 306; Conservative 304; Mismatches 663; Indels 330; Gaps 58;

```
Qy     43 ADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR 102
Db    738 AEAERQRRLDYES-QTAHDNLLTEQ--IHSLSIEAKSKDVKI---EVLQNELDDVQ---- 787

Qy    103 VEEMAANSTAVEDITKDGQEETSEII---EQIPASENNVEEMVQPAESQANDVGFKKVFK 159
Db    788 -LQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEELSQALSQKELEI------- 839

Qy    160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESEL 219
Db    840 -------TKMD--------QLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEE 884

Qy    220 KQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSE 279
Db    885 KFSLGVEIKTLKE-----QLNLLSRAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKE 939

Qy    280 TTSSFKKFFTHGWAGWRKKTSFKKSK--------EDDLETAEKRKEQEAEKVDEEEKEKT 331
Db    940 E-------LQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIP 992

Qy    332 EPASEEQEPAEDTDQARLS-----ADYEKVELPLEDQVGDLEA-------SSEEKCAPLA 379
```

```
Db     993 LSETERGEVEEDKENKEYSEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEK---LA 1049

Qy     380 TEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEV 439
Db    1050 AE------EQFQALVKQMNQTLQDKTNQ-----------------IDLLQAEISENQAI 1085

Qy     440 PQEAEPAEELMKSREMCVSGGDHTQLTD----LSPEEKTLPKHPEGIVSEVEMLSSQERI 495
Db    1086 IQK-------LITSNTDASDGDSVALVKETVVISPPCTGSSEHWKPELEEKILALEKEKE 1138

Qy     496 KVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGES--- 552
Db    1139 QLQKKLQEALTSRKAILKKAQEKERHLR-----EELKQQKDDYNRLQEQFDEQSKENENI 1193

Qy     553 ---------------SASSPEEPEETTCLEKGPLEAPQDGEAEEGTT------------- 584
Db    1194 GDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDWP 1253

Qy     585 --SDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK-SATLSSTDSTVSEMQ 641
Db    1254 SHSEDASALQGGTSVAQIKAQL-------------KEIEAEKVELELKVSSTTSELTKKS 1300

Qy     642 DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRA 701
Db    1301 EEVFQLQEQINKQGLEIESLKTVSHEAEVHAESLQQKL--------ESSQLQIAGLEHLR 1352

Qy     702 EEASKDKEAGTDAVPASTQEQDQAQGS---SSPEPAGSPSEGEGVSTWESFKRLVTPRKK 758
Db    1353 ELQPKLDEL---QKLISKKEEDVSYLSGQLSEKEAALTKIQTEIIEQEDLIKALHT---- 1405

Qy     759 SKSKLEEKAEDSSVEQLSTEI-------EPSREESWVSIKKFIPGRRKKRADGKQEQATV 811
Db    1406 -QLEMQAKEHDERIKQLQVELCEMKQKPEEIGEES----------RAKQQIQRKLQAALI 1454

Qy     812 EDSGPVEINEDDPNVPAVVPLSEYNAVER--------EKMEAQGNTELPQLLGAVYVSEE 863
Db    1455 SRK---EALKENKSLQEELSLAR-GTIERLTKSLADVESQVSAQNKEKDTVLGRLALLQE 1510

Qy     864 LSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA 923
Db    1511 ERDKLITEMDRSLLENQSLSSSCE--SLKLALEGLTEDKEKLVKEI-----ESLKSSKIA 1563

Qy     924 EETPVLTQTLPEGKDAHDD-------MVTSEVDFTSEAVTATETSEALRTEE-------- 968
Db    1564 EST--------EWQEKHKELQKEYEILLQSYENVSNEAERIQHVVEAVRQEKQELYGKLR 1615

Qy     969 VTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADK 1028
Db    1616 STEANKKETEKQLQEAEQEMEEMKEKMRKF--AKSKQQKILELEEENDRLRAEVHPAGDT 1673

Qy    1029 VKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASE-----KEKDVMPKGPVQEAGAEH 1083
Db    1674 AKE-----CMETLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSE-EVQD--LKH 1725

Qy    1084 LAQGSETGQATPESLEVPEVTADVDHVATCQV---IKLQQLMEQAVAPESSETLTDSETN 1140
Db    1726 QIEGNVSKQANLEATEKHDNQTNVTEEGTQSIPGETEEQDSLSMSTRPTCSESVPSAKS- 1784

Qy    1141 GSTPLADSDTADGTQQD---ETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVP 1197
Db    1785 -ANPAVSKDFSSHDEINNYLQQIDQLKERIAGLEEEKQKNKEFSQTLENEK-NTLLSQIS 1842

Qy    1198 AQEEHGEEPGRDVLEPTQQELT---------AAAVPVLAKTEVGQEGEVDWLDGEKVKEE 1248
Db    1843 TKD--GE------LKMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKDDLEERLMNQL 1894

Qy    1249 QEVFVHSGPNSQKAADVT--------YDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVE 1300
Db    1895 AEL---NGSIGNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVKEKTKVESEIR 1951

Qy    1301 KEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDK 1360
Db    1952 KEYLEKIQGAQKEPGNK---------SHAKELQELLKEKQQEVKQL----QKDCIRYQEK 1998

Qy    1361 AGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH------LLPAEKSSAT 1414
Db    1999 ISALERTVKALEFVQTESQKDLEITKEN--LAQAVEHRKKAQAELASFKVLLDDTQSEAA 2056

Qy    1415 GGHWTLQHAEDTVPLGPESQA--ESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKP 1472
Db    2057 ------RVLADNLKLKKELQSNKESVKSQMKQKDE-----DLERRLEQAEEKHLKEKKNM 2105

Qy    1473 DAGPDADGKESTAIEKVL--------KAEPEILELESKSNKIV 1507
Db    2106 QEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQENLDSTV 2148
```

RESULT 15
US-09-724-676A-55849
; Sequence 55849, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 55849
;   LENGTH: 2915
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-724-676A-55849

  Query Match   5.0%; Score 400; DB 5; Length 2915;
  Best Local Similarity   19.1%; Pred. No. 7.9e-15;
  Matches 306; Conservative 304; Mismatches 663; Indels 330; Gaps 58;

```
Qy      43 ADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR 102
Db     738 AEAERQRRLDYES-QTAHDNLLTEQ--IHSLSIEAKSKDVKI---EVLQNELDDVQ---- 787

Qy     103 VEEMAANSTAVEDITKDGQEETSEII---EQIPASENNVEEMVQPAESQANDVGFKKVFK 159
Db     788 -LQFSEQSTLIRSLQSQLQNKESEVLEGAERVRHISSKVEELSQALSQKELEI------- 839

Qy     160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESEL 219
Db     840 -------TKMD--------QLLLEKKRDVETLQQTIEEKDQQVTEISFSMTEKMVQLNEE 884

Qy     220 KQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSE 279
Db     885 KFSLGVEIKTLKE-----QLNLLSRAEEAKKEQVEEDNEVSSGLKQNYDEMSPAGQISKE 939

Qy     280 TTSSFKKFFTHGWAGWRKKTSFKKSK--------EDDLETAEKRKEQEAEKVDEEEKEKT 331
Db     940 E-------LQHEFDLLKKENEQRKRKLQAALINRKELLQRVSRLEEELANLKDESKKEIP 992

Qy     332 EPASEEQEPAEDTDQARLS-----ADYEKVELPLEDQVGDLEA-------SSEEKCAPLA 379
Db     993 LSETERGEVEEDKENKEYSEKCVTSKCQEIEIYLKQTISEKEVELQHIRKDLEEK---LA 1049

Qy     380 TEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEV 439
Db    1050 AE------EQFQALVKQMNQTLQDKTNQ-----------------IDLLQAEISENQAI 1085

Qy     440 PQEAEPAEELMKSREMCVSGGDHTQLTD----LSPEEKTLPKHPEGIVSEVEMLSSQERI 495
Db    1086 IQK-------LITSNTDASDGDSVALVKETVVISPPCTGSSEHWKPELEEKILALEKEKE 1138

Qy     496 KVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGES--- 552
Db    1139 QLQKKLQEALTSRKAILKKAQEKERHLR-----EELKQQKDDYNRLQEQFDEQSKENENI 1193

Qy     553 ---------------SASSPEEPEETTCLEKGPLEAPQDGEAEEGTT------------- 584
```

```
Db   1194 GDQLRQLQIQVRESIDGKLPSTDQQESCSSTPGLEEPLFKATEQHHTQPVLESNLCPDWP 1253

Qy    585 --SDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK-SATLSSTDSTVSEMQ 641
Db   1254 SHSEDASALQGGTSVAQIKAQL------------KEIEAEKVELELKVSSTTSELTKKS 1300

Qy    642 DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRA 701
Db   1301 EEVFQLQEQINKQGLEIESLKTVSHEAEVHAESLQQKL--------ESSQLQIAGLEHLR 1352

Qy    702 EEASKDKEAGTDAVPASTQEQDQAQGS---SSPEPAGSPSEGEGVSTWESFKRLVTPRKK 758
Db   1353 ELQPKLDEL---QKLISKKEEDVSYLSGQLSEKEAALTKIQTEIIEQEDLIKALHT---- 1405

Qy    759 SKSKLEEKAEDSSVEQLSTEI-------EPSREESWVSIKKFIPGRRKKRADGKQEQATV 811
Db   1406 -QLEMQAKEHDERIKQLQVELCEMKQKPEEIGEES----------RAKQQIQRKLQAALI 1454

Qy    812 EDSGPVEINEDDPNVPAVVPLSEYNAVER--------EKMEAQGNTELPQLLGAVYVSEE 863
Db   1455 SRK---EALKENKSLQEELSLAR-GTIERLTKSLADVESQVSAQNKEKDTVLGRLALLQE 1510

Qy    864 LSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA 923
Db   1511 ERDKLITEMDRSLLENQSLSSSCE--SLKLALEGLTEDKEKLVKEI-----ESLKSSKIA 1563

Qy    924 EETPVLTQTLPEGKDAHDD-------MVTSEVDFTSEAVTATETSEALRTEE-------- 968
Db   1564 EST--------EWQEKHKELQKEYEILLQSYENVSNEAERIQHVVEAVRQEKQELYGKLR 1615

Qy    969 VTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADK 1028
Db   1616 STEANKKETEKQLQEAEQEMEEMKEKMRKF--AKSKQQKILELEEENDRLRAEVHPAGDT 1673

Qy   1029 VKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASE-----KEKDVMPKGPVQEAGAEH 1083
Db   1674 AKE-----CMETLLSSNASMKEELERVKMEYETLSKKFQSLMSEKDSLSE-EVQD--LKH 1725

Qy   1084 LAQGSETGQATPESLEVPEVTADVDHVATCQV---IKLQQLMEQAVAPESSETLTDSETN 1140
Db   1726 QIEGNVSKQANLEATEKHDNQTNVTEEGTQSIPGETEEQDSLSMSTRPTCSESVPSAKS- 1784

Qy   1141 GSTPLADSDTADGTQQD---ETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVP 1197
Db   1785 -ANPAVSKDFSSHDEINNYLQQIDQLKERIAGLEEEKQKNKEFSQTLENEK-NTLLSQIS 1842

Qy   1198 AQEEHGEEPGRDVLEPTQQELT---------AAAVPVLAKTEVGQEGEVDWLDGEKVKEE 1248
Db   1843 TKD--GE------LKMLQEEVTKMNLLNQQIQEELSRVTKLKETAEEEKDDLEERLMNQL 1894

Qy   1249 QEVFVHSGPNSQKAADVT--------YDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVE 1300
Db   1895 AEL---NGSIGNYCQDVTDAQIKNELLESEMKNLKKCVSELEEEKQQLVKEKTKVESEIR 1951

Qy   1301 KEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDK 1360
Db   1952 KEYLEKIQGAQKEPGNK---------SHAKELQELLKEKQQEVKQL----QKDCIRYQEK 1998

Qy   1361 AGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH------LLPAEKSSAT 1414
Db   1999 ISALERTVKALEFVQTESQKDLEITKEN--LAQAVEHRKKAQAELASFKVLLDDTQSEAA 2056

Qy   1415 GGHWTLQHAEDTVPLGPESQA--ESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKP 1472
Db   2057 ------RVLADNLKLKKELQSNKESVKSQMKQKDE-----DLERRLEQAEEKHLKEKKNM 2105

Qy   1473 DAGPDADGKESTAIEKVL--------KAEPEILELESKSNKIV 1507
Db   2106 QEKLDALRREKVHLEETIGEIQVTLNKKDKEVQQLQENLDSTV 2148
```

Search completed: December 13, 2002, 00:24:39
Job time : 110 secs

```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         December 13, 2002, 00:28:41 ; Search time 6461 Seconds
                                  (without alignments)
                                  6210.712 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters:      49582208

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09902432/runat_12122002_141453_25861/app_query.fasta_1.1735
-DB=Pending_Patents_NA_Main -QFMT=fastap -SUFFIX=rnpm -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09902432_@CGN_1_1_5452_@runat_12122002_141453_25861 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Pending_Patents_NA_Main:*
                1:  /cgn2_6/ptodata/2/pna/PCTUS_COMB.seq:*
                2:  /cgn2_6/ptodata/2/pna/US06_COMB.seq:*
                3:  /cgn2_6/ptodata/2/pna/US07_COMB.seq:*
                4:  /cgn2_6/ptodata/2/pna/US080_COMB.seq:*
                5:  /cgn2_6/ptodata/2/pna/US081_COMB.seq:*
                6:  /cgn2_6/ptodata/2/pna/US082_COMB.seq:*
                7:  /cgn2_6/ptodata/2/pna/US083_COMB.seq:*
                8:  /cgn2_6/ptodata/2/pna/US084_COMB.seq:*
                9:  /cgn2_6/ptodata/2/pna/US085_COMB.seq:*
                10:  /cgn2_6/ptodata/2/pna/US086_COMB.seq:*
                11:  /cgn2_6/ptodata/2/pna/US087_COMB.seq:*
                12:  /cgn2_6/ptodata/2/pna/US088_COMB.seq:*
                13:  /cgn2_6/ptodata/2/pna/US089_COMB.seq:*
                14:  /cgn2_6/ptodata/2/pna/US090_COMB.seq:*
                15:  /cgn2_6/ptodata/2/pna/US091_COMB.seq:*
                16:  /cgn2_6/ptodata/2/pna/US092_COMB.seq:*
                17:  /cgn2_6/ptodata/2/pna/US093_COMB.seq:*
                18:  /cgn2_6/ptodata/2/pna/US094_COMB.seq:*
                19:  /cgn2_6/ptodata/2/pna/US095A_COMB.seq:*
                20:  /cgn2_6/ptodata/2/pna/US095B_COMB.seq:*
                21:  /cgn2_6/ptodata/2/pna/US095C_COMB.seq:*
                22:  /cgn2_6/ptodata/2/pna/US095D_COMB.seq:*
                23:  /cgn2_6/ptodata/2/pna/US096A_COMB.seq:*
                24:  /cgn2_6/ptodata/2/pna/US096B_COMB.seq:*
                25:  /cgn2_6/ptodata/2/pna/US096C_COMB.seq:*
                26:  /cgn2_6/ptodata/2/pna/US096D_COMB.seq:*
                27:  /cgn2_6/ptodata/2/pna/US096E_COMB.seq:*
                28:  /cgn2_6/ptodata/2/pna/US097A_COMB.seq:*
                29:  /cgn2_6/ptodata/2/pna/US097B_COMB.seq:*
                30:  /cgn2_6/ptodata/2/pna/US097C_COMB.seq:*
                31:  /cgn2_6/ptodata/2/pna/US098A_COMB.seq:*
                32:  /cgn2_6/ptodata/2/pna/US098B_COMB.seq:*
                33:  /cgn2_6/ptodata/2/pna/US098C_COMB.seq:*
                34:  /cgn2_6/ptodata/2/pna/US099A_COMB.seq:*
                35:  /cgn2_6/ptodata/2/pna/US099B_COMB.seq:*
                36:  /cgn2_6/ptodata/2/pna/US099C_COMB.seq:*
                37:  /cgn2_6/ptodata/2/pna/US099D_COMB.seq:*
                38:  /cgn2_6/ptodata/2/pna/US100A_COMB.seq:*
                39:  /cgn2_6/ptodata/2/pna/US100B_COMB.seq:*
                40:  /cgn2_6/ptodata/2/pna/US101A_COMB.seq:*
                41:  /cgn2_6/ptodata/2/pna/US101B_COMB.seq:*
                42:  /cgn2_6/ptodata/2/pna/US102A_COMB.seq:*
                43:  /cgn2_6/ptodata/2/pna/US102B_COMB.seq:*
                44:  /cgn2_6/ptodata/2/pna/US6000_COMB.seq:*
                45:  /cgn2_6/ptodata/2/pna/US6001_COMB.seq:*
                46:  /cgn2_6/ptodata/2/pna/US6002_COMB.seq:*
                47:  /cgn2_6/ptodata/2/pna/US6003_COMB.seq:*
                48:  /cgn2_6/ptodata/2/pna/US6004_COMB.seq:*
                49:  /cgn2_6/ptodata/2/pna/US6005_COMB.seq:*
                50:  /cgn2_6/ptodata/2/pna/US6006_COMB.seq:*
                51:  /cgn2_6/ptodata/2/pna/US6007_COMB.seq:*
                52:  /cgn2_6/ptodata/2/pna/US6008_COMB.seq:*
                53:  /cgn2_6/ptodata/2/pna/US6009_COMB.seq:*
                54:  /cgn2_6/ptodata/2/pna/US6010_COMB.seq:*
                55:  /cgn2_6/ptodata/2/pna/US6011_COMB.seq:*
                56:  /cgn2_6/ptodata/2/pna/US6012_COMB.seq:*
                57:  /cgn2_6/ptodata/2/pna/US6013_COMB.seq:*
                58:  /cgn2_6/ptodata/2/pna/US6014_COMB.seq:*
                59:  /cgn2_6/ptodata/2/pna/US6015_COMB.seq:*
                60:  /cgn2_6/ptodata/2/pna/US6016_COMB.seq:*
                61:  /cgn2_6/ptodata/2/pna/US6017_COMB.seq:*
                62:  /cgn2_6/ptodata/2/pna/US6018_COMB.seq:*
                63:  /cgn2_6/ptodata/2/pna/US6019_COMB.seq:*
                64:  /cgn2_6/ptodata/2/pna/US6020_COMB.seq:*
                65:  /cgn2_6/ptodata/2/pna/US6021_COMB.seq:*
                66:  /cgn2_6/ptodata/2/pna/US6022_COMB.seq:*
                67:  /cgn2_6/ptodata/2/pna/US6023_COMB.seq:*
                68:  /cgn2_6/ptodata/2/pna/US6024_COMB.seq:*
                69:  /cgn2_6/ptodata/2/pna/US6025_COMB.seq:*
                70:  /cgn2_6/ptodata/2/pna/US6026_COMB.seq:*
                71:  /cgn2_6/ptodata/2/pna/US6027_COMB.seq:*
                72:  /cgn2_6/ptodata/2/pna/US6028_COMB.seq:*
                73:  /cgn2_6/ptodata/2/pna/US6029_COMB.seq:*
                74:  /cgn2_6/ptodata/2/pna/US6030_COMB.seq:*
                75:  /cgn2_6/ptodata/2/pna/US6031_COMB.seq:*
                76:  /cgn2_6/ptodata/2/pna/US6032_COMB.seq:*
                77:  /cgn2_6/ptodata/2/pna/US6033_COMB.seq:*
                78:  /cgn2_6/ptodata/2/pna/US6034_COMB.seq:*
                79:  /cgn2_6/ptodata/2/pna/US6035_COMB.seq:*
                80:  /cgn2_6/ptodata/2/pna/US6036_COMB.seq:*
                81:  /cgn2_6/ptodata/2/pna/US6037_COMB.seq:*
                82:  /cgn2_6/ptodata/2/pna/US6038_COMB.seq:*
                83:  /cgn2_6/ptodata/2/pna/US6039_COMB.seq:*
                84:  /cgn2_6/ptodata/2/pna/US6040_COMB.seq:*
                85:  /cgn2_6/ptodata/2/pna/US6041_COMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8073	100.0	5200	13	US-08-978-277A-3	Sequence 3, Appli
2	8073	100.0	6160	1	PCT-US97-06830-3	Sequence 3, Appli
3	8073	100.0	6160	10	US-08-665-401-3	Sequence 3, Appli
4	8073	100.0	6160	34	US-09-902-432-3	Sequence 3, Appli
5	8030	99.5	5236	1	PCT-US02-16173A-2063	Sequence 2063, Ap
6	8030	99.5	5236	41	US-10-191-803-279	Sequence 279, App

```
   7    6663   82.5    6199  80  US-60-360-207-12210       Sequence 12210, A
   8    6005   74.4    5074  13  US-08-978-277A-1          Sequence 1, Appli
   9    6005   74.4    5134   1  PCT-US97-06830-1          Sequence 1, Appli
  10    6005   74.4    5134  10  US-08-665-401-1           Sequence 1, Appli
  11    6005   74.4    5134  34  US-09-902-432-1           Sequence 1, Appli
  12  4334.5   53.7    7651  30  US-09-770-174-4018        Sequence 4018, Ap
  13  4333.5   53.7    6603  71  US-60-278-232-10143       Sequence 10143, A
c 14  4325.5   53.6    6607  68  US-60-245-228-517         Sequence 517, App
  15    4316   53.5    8660  25  US-09-644-867-6364        Sequence 6364, Ap
  16    4316   53.5    8660  25  US-09-644-869-9117        Sequence 9117, Ap
  17    4316   53.5    8660  25  US-09-652-911-10117       Sequence 10117, A
  18    4316   53.5    8660  25  US-09-652-913-10715       Sequence 10715, A
  19    4316   53.5    8660  25  US-09-652-918-8758        Sequence 8758, Ap
  20    4316   53.5    8660  26  US-09-667-760-2000        Sequence 2000, Ap
  21    4316   53.5    8660  27  US-09-699-997-11667       Sequence 11667, A
  22    4316   53.5    8660  27  US-09-699-999-6654        Sequence 6654, Ap
  23    4316   53.5    8660  28  US-09-700-000-5679        Sequence 5679, Ap
  24    4316   53.5    8660  28  US-09-710-280-2315        Sequence 2315, Ap
  25    4316   53.5    8660  28  US-09-710-286-3428        Sequence 3428, Ap
  26    4316   53.5    8660  28  US-09-716-472-4599        Sequence 4599, Ap
  27    4316   53.5    8660  28  US-09-716-953-2245        Sequence 2245, Ap
  28    4316   53.5    8660  29  US-09-726-789-3701        Sequence 3701, Ap
  29    4316   53.5    8660  29  US-09-726-802-2407        Sequence 2407, Ap
  30    4316   53.5    8660  29  US-09-726-805-1797        Sequence 1797, Ap
  31    4316   53.5    8660  29  US-09-726-806-5392        Sequence 5392, Ap
  32    4311   53.4    8492  23  US-09-606-776-5120        Sequence 5120, Ap
  33  4310.5   53.4    6632  76  US-60-324-185-30580       Sequence 30580, A
  34  4300.5   53.3    5346  29  US-09-738-877-2           Sequence 2, Appli
  35  4300.5   53.3    6608  16  US-09-220-132-58          Sequence 58, Appl
  36  4300.5   53.3    6608  29  US-09-738-877-1           Sequence 1, Appli
  37  4300.5   53.3    6608  33  US-09-880-107-3439        Sequence 3439, Ap
  38  4300.5   53.3    6608  41  US-10-155-653-58          Sequence 58, Appl
  39  4226.5   52.4    6886   1  PCT-US01-08631-21009      Sequence 21009, A
  40  4106.5   50.9  120896  68  US-60-245-228-97          Sequence 97, Appl
  41  4104.5   50.8    8886  76  US-60-324-185-30584       Sequence 30584, A
  42  4102.5   50.8  174127  65  US-60-212-664-282         Sequence 282, App
  43  4096.5   50.7    8877  61  US-60-172-360-24882       Sequence 24882, A
  44  4091.5   50.7    8869  71  US-60-278-232-10145       Sequence 10145, A
  45  4091.5   50.7    8888  65  US-60-213-359-7019        Sequence 7019, Ap
```

ALIGNMENTS

```
RESULT 1
US-08-978-277A-3
; Sequence 3, Application US/08978277A
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,277A
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5200 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-978-277A-3

Alignment Scores:
Pred. No.:                0              Length:        5200
Score:                    8073.00        Matches:       1596
Percent Similarity:       100.00%        Conservative:  0
Best Local Similarity:    100.00%        Mismatches:    0
Query Match:              100.00%        Indels:        0
DB:                       13             Gaps:          0

US-09-902-432-4 (1-1596) x US-08-978-277A-3 (1-5200)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 60

Qy     21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 120

Qy     41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 180

Qy     61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 240

Qy     81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 300

Qy    101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 360

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 420

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 480

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 540

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 600

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 660

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     661 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 720

Qy     241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 780

Qy     261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 840

Qy     281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 900

Qy     301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 960

Qy     321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1020

Qy     341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1080

Qy     361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1140

Qy     381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1200

Qy     401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1260

Qy     421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1320

Qy     441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1380

Qy     461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1440

Qy     481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1500

Qy     501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1560

Qy     521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1561 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1620

Qy     541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1621 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1680

Qy     561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1681 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1740

Qy     581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1741 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1800

Qy     601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1801 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1860

Qy     621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1861 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1920

Qy     641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1921 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 1980

Qy     661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1981 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2040

Qy     681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2041 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2100

Qy     701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2101 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2160

Qy     721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2161 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2220

Qy     741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2221 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2280

Qy     761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2281 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2340

Qy     781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2341 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2400

Qy     801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2401 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2460

Qy     821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2461 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2520

Qy     841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2521 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2580

Qy     861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2581 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2640

Qy     881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2641 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2700

Qy     901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2701 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2760

Qy     921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2761 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2820

Qy     941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2821 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2880
```

```
Qy    961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2940

Qy    981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3000

Qy   1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluArgGlnThrGlnAla 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3060

Qy   1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3120

Qy   1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3180

Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3240

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3300

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3360

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3420

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3480

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3540

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3600

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3660

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3720

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3780

Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3840

Qy   1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3900

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3960

Qy   1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4020

Qy   1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4080

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4140

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4200

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4260

Qy   1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4320

Qy   1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4380

Qy   1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4440

Qy   1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4500

Qy   1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4560

Qy   1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4620

Qy   1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4680

Qy   1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4740

Qy   1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4788
```

```
RESULT 2
PCT-US97-06830-3
; Sequence 3, Application PC/TUS9706830
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
```

```
;     SOFTWARE:  FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/US97/06830
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  08/665,401
;     FILING DATE:  18-JUN-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  Clark, Richard S
;     REGISTRATION NUMBER:  26,154
;     REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  212-408-2558
;     TELEFAX:  212-765-2519
;     TELEX:
; INFORMATION FOR SEQ ID NO:  3:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH:  6160 base pairs
;     TYPE:  nucleic acid
;     STRANDEDNESS:  single
;     TOPOLOGY:  linear
;   MOLECULE TYPE:  cDNA
;   HYPOTHETICAL:  NO
;   ANTI-SENSE:  NO
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:
PCT-US97-06830-3

Alignment Scores:
Pred. No.:              0              Length:        6160
Score:                  8073.00        Matches:       1596
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     1              Gaps:          0

US-09-902-432-4 (1-1596) x PCT-US97-06830-3 (1-6160)

Qy       1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92

Qy      21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy      41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy      61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy      81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy     101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392

Qy     121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy     141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy     161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy     181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy     201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy     221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy     241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy     261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy     281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy     301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy     321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy     341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy     361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy     381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy     401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy     421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy     441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy     461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy     481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532

Qy     501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592

Qy     521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652
```

```
Qy    541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy    561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy    581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy    601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy    621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy    641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy    661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy    681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132

Qy    701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy    741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy    761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2433 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2493 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2552

Qy    841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2553 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612

Qy    861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy    881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy    901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy    921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy    941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy    961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy    981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy   1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy   1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy   1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212

Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812

Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy      1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy      1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy      1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy      1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112

Qy      1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy      1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy      1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy      1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy      1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy      1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy      1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy      1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy      1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4652

Qy      1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4712

Qy      1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772

Qy      1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db      4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820

RESULT 3
US-08-665-401-3
; Sequence 3, Application US/08665401
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/665,401
;      FILING DATE:  18-JUN-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6160 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-665-401-3

Alignment Scores:
Pred. No.:              0               Length:         6160
Score:                  8073.00         Matches:        1596
Percent Similarity:     100.00%         Conservative:   0
Best Local Similarity:  100.00%         Mismatches:     0
Query Match:            100.00%         Indels:         0
DB:                     10              Gaps:           0

US-09-902-432-4 (1-1596) x US-08-665-401-3 (1-6160)

Qy         1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGAGCGACACG 92

Qy        21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy        41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy        61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy        81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy       101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy    341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy    361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy    381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy    401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy    421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy    441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy    461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy    481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532

Qy    501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592

Qy    521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652

Qy    541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy    561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy    581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy    601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy    621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy    641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy    661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy    681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132

Qy    701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy    741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy    761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2433 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2493 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2552
```

```
Qy    841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2553 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612

Qy    861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy    881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy    901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy    921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy    941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy    961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy    981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy   1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy   1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy   1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212

Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812

Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy   1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy   1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy   1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy   1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy   1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy   1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy   1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy   1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy   1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4652

Qy   1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4712

Qy   1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772

Qy   1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820

RESULT 4
US-09-902-432-3
; Sequence 3, Application US/09902432
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 6160
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-902-432-3

Alignment Scores:
Pred. No.:               0            Length:        6160
Score:                   8073.00      Matches:       1596
Percent Similarity:      100.00%      Conservative:  0
Best Local Similarity:   100.00%      Mismatches:    0
Query Match:             100.00%      Indels:        0
DB:                      34           Gaps:          0

US-09-902-432-4 (1-1596) x US-09-902-432-3 (1-6160)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92

Qy     21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy     41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy     61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy     81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy    101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy    341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy    361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy    381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy    401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy    421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy    441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy    461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy    481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532

Qy    501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592

Qy    521 GlyLysArgGlyGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1593 GGGAAACGAGGAGGTGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652

Qy    541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy    561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy    581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy    601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy    621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy    641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy    661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy    681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132

Qy    701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy    741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy    761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2433 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2493 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2552

Qy    841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2553 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612

Qy    861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy    881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy    901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy    921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy    941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy    961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy    981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy   1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy   1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy   1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212

Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812
```

```
Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy   1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy   1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy   1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy   1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy   1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy   1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy   1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy   1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy   1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4652

Qy   1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4712

Qy   1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772

Qy   1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820

RESULT 5
PCT-US02-16173A-2063
; Sequence 2063, Application PC/TUS0216173A
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Higgs, Brandon
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5089-WO
;  CURRENT APPLICATION NUMBER: PCT/US02/16173A
;  CURRENT FILING DATE:  2002-10-02
;  PRIOR APPLICATION NUMBER: US 60/292,335
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/297,523
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,925
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,810
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,807
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,808
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/315,047
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/324,928
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/330,867
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/330,462
;  PRIOR FILING DATE: 2001-10-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
;  NUMBER OF SEQ ID NOS: 2221
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2063
;   LENGTH: 5236
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. NM_057103
PCT-US02-16173A-2063

Alignment Scores:
Pred. No.:              0          Length:        5236
Score:                  8030.00    Matches:       1589
Percent Similarity:     99.75%     Conservative:  3
Best Local Similarity:  99.56%     Mismatches:    4
Query Match:            99.47%     Indels:        0
DB:                     1          Gaps:          0

US-09-902-432-4 (1-1596) x PCT-US02-16173A-2063 (1-5236)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92

Qy     21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy     41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy     61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy     81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy    101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGG 392
```

```
Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy    341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy    361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy    381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy    401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy    421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy    441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy    461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy    481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532

Qy    501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592

Qy    521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652

Qy    541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy    561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy    581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy    601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy    621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy    641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy    661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy    681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132

Qy    701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy    741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy    761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2433 AGGGCAGATGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
          ||||||||||||:::||||||||||||||||||||||||||||||:::||||||||||||
Db   2493 GAGGACGACCCTGATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGG 2552

Qy    841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
```

```
                 ||||||||||||||||||||||      ||||||||||||||||||||||||||||||||
Db     2553 GAGAAGATGGAAGCCCAGGGGAATGCGGAGCTGCCCCAGCTGCTGGGGCTGTGTACGTG 2612

Qy      861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy      881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy      901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy      921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy      941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy      961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy      981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy     1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluArgGlnThrGlnAla 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy     1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy     1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212

Qy     1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy     1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy     1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy     1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy     1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy     1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy     1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy     1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy     1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy     1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812

Qy     1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy     1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy     1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3933 AAGGAGAAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy     1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy     1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112

Qy     1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy     1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy     1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy     1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy     1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy     1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy     1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy     1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy     1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
            |||||||||||||||||||||||||||||||||||||||||||||||||     :::
Db     4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCAG 4652

Qy     1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
               ||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db     4653 GCTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACAC 4712

Qy     1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772
```

```
Qy   1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820

RESULT 6
US-10-191-803-279
; Sequence 279, Application US/10191803
; GENERAL INFORMATION:
;  APPLICANT: MENDRICK, Donna
;  APPLICANT:  PORTER, Mark
;  APPLICANT:  JOHNSON, Kory
;  APPLICANT:  HIGGS, Brandon
;  APPLICANT:  CASTLE, Arthur
;  APPLICANT:  ELASHOFF, Michael
;  TITLE OF INVENTION: Cardiotoxin Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5090US
;  CURRENT APPLICATION NUMBER: US/10/191,803
;  CURRENT FILING DATE:  2002-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,819
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/305,623
;  PRIOR FILING DATE: 2001-07-17
;  PRIOR APPLICATION NUMBER: US 60/369,351
;  PRIOR FILING DATE: 2002-04-03
;  PRIOR APPLICATION NUMBER: US 60/377,611
;  PRIOR FILING DATE: 2002-05-06
;  NUMBER OF SEQ ID NOS: 1140
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 279
;   LENGTH: 5236
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. NM_057103
US-10-191-803-279

Alignment Scores:
Pred. No.:              0             Length:         5236
Score:                  8030.00       Matches:        1589
Percent Similarity:     99.75%        Conservative:   3
Best Local Similarity:  99.56%        Mismatches:     4
Query Match:            99.47%        Indels:         0
DB:                     41            Gaps:           0

US-09-902-432-4 (1-1596) x US-10-191-803-279 (1-5236)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92

Qy     21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy     41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy     61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy     81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy    101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy    341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy    361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy    381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy    401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy    421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy    441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy    461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy    481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532
```

```
Qy      501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592

Qy      521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652

Qy      541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy      561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy      581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy      601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy      621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy      641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy      661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy      681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132

Qy      701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy      721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy      741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy      761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy      781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy      801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2433 AGGGCAGATGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy      821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
            ||||||||||||:::|||||||||||||||||||||||||||||||||:::|||||||||
Db     2493 GAGGACGACCCTGATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGG 2552

Qy      841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
            ||||||||||||||||||||||||   |||||||||||||||||||||||||||||||||
Db     2553 GAGAAGATGGAAGCCCAGGGGAATGCGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612

Qy      861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy      881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy      901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy      921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy      941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy      961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy      981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy     1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy     1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy     1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212

Qy     1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy     1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy     1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy     1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy     1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy     1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy     1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy     1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy     1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
```

```
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy    1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812

Qy    1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy    1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy    1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy    1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy    1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112

Qy    1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy    1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy    1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy    1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy    1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy    1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy    1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy    1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy    1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||   :::
Db    4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCAG 4652

Qy    1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
              ||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db    4653 GCTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACAC 4712

Qy    1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772

Qy    1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
           ||||||||||||||||||||||||||||||||||||||||||||||||

Db    4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820
```

RESULT 7
US-60-360-207-12210
; Sequence 12210, Application US/60360207
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig
; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/60/360,207
; CURRENT FILING DATE: 2002-03-01
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 12210
;   LENGTH: 6199
;   TYPE: DNA
;   ORGANISM: HUMAN
US-60-360-207-12210

Alignment Scores:

Pred. No.:	0	Length:	6199
Score:	6663.00	Matches:	1357
Percent Similarity:	88.90%	Conservative:	76
Best Local Similarity:	84.18%	Mismatches:	156
Query Match:	82.53%	Indels:	23
DB:	80	Gaps:	11

US-09-902-432-4 (1-1596) x US-60-360-207-12210 (1-6199)

```
Qy       1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
           |||||||||||||||||||||||||||||||||||||||||||||||    |||||||||
Db     142 ATGGGTGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGAGAGCGACACG 201

Qy      21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
           ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db     202 CCGAGCGAGCTGGAGCTCAGTGGCCATGGGCCCGCAGCGGAAGCGTCGGGAGCAGCTGGA 261

Qy      41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db     262 GATCCCGCTGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGTCAGCTGTCTGCC 321

Qy      61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
           ||||||||||||||||||||    ||||||||||||||||||:::|||:::||||||
Db     322 GTCAATGGTGTAGCTGAACAAGAAGATGTCCACGTCCAAGAGGAAAGCCAGGATGGGCAA 381

Qy      81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
           ||||||||||||      ||||||||||||||||||||||||||||||:::||||||
Db     382 GAGGAAGAAGTCACTGTTGAAGATGTTGGACAGAGAGAGTCAGAAGATGTGAAAGAAAAA 441

Qy     101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
           |||||    :::|||||||||||:::|||||||   |||||||||||||||||||||
Db     442 GACCGAGCTAAAGAAATGGCAGCCAGTTCCACAGTTGTTGAAGATATCACAAAGGACGAG 501

Qy     121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
           ||||||||||||  ||||||||||||||||||||||||||||:::|||||||||||||
Db     502 CAGGAGGAAACACCGGAAATAATCGAACAGATCCCTGCTTCAGAGAGCAATGTGGAAGAA 561

Qy     141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
           |||   |||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db     562 ATGGCGCAGGCTGCTGAGTCCCAAGCTAATGACGTCGGCTTCAAGAAGGTATTTAAATTT 621

Qy     161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     622 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAACGAAAAGTCAGATACCGTCCAGCTA 681

Qy     181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     682 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCCGTCGGAGCAGGAGACCAC 741

Qy     201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
           |||||||||   ||||||||   |||||||||||||||||||||||||||||||||||||
Db     742 CAAGAGCCCGGAGTGGAGACC---GTCGGCGAATCAGCATCCAAAGAAAGTGAGCTGAAG 798

Qy     221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
```

```
Db    799 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGCACAGAGCAGCACAGAAATTCCC 858

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
Db    859 CTTCAAGCCGAATCTGGTCAAGGGACCGAGGAAGAAGCAGCCAAAGATGGAGAAGAAAAC 918

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
Db    919 CGAGAGAAAGAACCTACCAAGCCCCTAGAATCTCCGACCAGCCCTGTCAGCAATGAGACA 978

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
Db    979 ACATCTTCCTTCAAGAAATTCTTCACTCACGGCTGGGCCGGCTGGCGCAAGAAGACCAGC 1038

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
Db   1039 TTCAAGAAACCAAAGGAAGATGATCTGGAAACTTCCGAGAAGAGAAAGGAGCAAGAGGCT 1098

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAla---SerGluGluGlnGlu 339
Db   1099 GAAAAAGTAGACGAGGAAGAAGGGGAAAAGACAGAGCCAGCCCCAGCCGAGGAGCAGGAG 1158

Qy    340 ProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuPro 359
Db   1159 CCTGCAGAAGGCACAGACCAGGCCAGGTTGTCAGCCGACTATGAGAAGGTGGAGTTGCCT 1218

Qy    360 LeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAla 379
Db   1219 TTGGAAGACCAGGTCGGTGACCTGGAGGCATTGTCG---GAGAAGTGTGCTCCTTTGGCA 1275

Qy    380 ThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisVal 399
Db   1276 ACGGAAGTGTTTGATGAGAAGACGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTG 1335

Qy    400 SerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyVal 419
Db   1336 AGCACCGTGGAGAAGATGACGAAAGGGCAA---GGAGGAGCAGAGGTGGAAGGGGATGTG 1392

Qy    420 ValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluVal 439
Db   1393 GTGGTGGAAGGATCGGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGACCCAGGAGGTC 1452

Qy    440 ProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGly 459
Db   1453 CCCCAGGAAGCTGAGCCTGTGGAGGAGCTGATGAAGACCAAAGAAGTATGCGTCTCTGGG 1512

Qy    460 GlyAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisPro 479
Db   1513 GGTGACCATACTCAGCTGACAGATCTAAGTCCTGAAGAGAAGATGCTACCCAAACACCCC 1572

Qy    480 GluGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGly 499
Db   1573 GAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAGAGAATCAAGGTACAGGGA 1632

Qy    500 SerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGln 519
Db   1633 AGTCCCCTGAAGAAGCTCTTCAGCAGTTCGGGCTTAAAGAAGCTCTCCGGGAAGAAGCAG 1692

Qy    520 LysGlyLysArg---GlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHis 538
Db   1693 AAGGGGAAGAGAGGAGGAGGCGGGGGAGATGAAGAGCCAGGAGAATACCAACACATTCAA 1752

Qy    539 ThrGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGlu 558
Db   1753 ACCGAGTCCCCAGAGAGTGCTGACGAGCAGAAGGGAGAGAGCTCTGCCTCTTCCCCTGAA 1812

Qy    559 GluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGlu 578
Db   1813 GAGCCCGAGGAGATCGCGTGTCTGGAGAAGGGGCCATCGGAAGCACCCCAGGAAGCGGAA 1872

Qy    579 AlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAla 598
Db   1873 GCTGAGGAAGGAGCGACTTCCGACGGAGAGAAGAAAAGGGAAGGGATCACCCCCTGGGCA 1932

Qy    599 SerPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLys 618
Db   1933 TCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGCGACAAG 1992

Qy    619 GluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSer 638
Db   1993 GAAGAAGAGCTGGATAAGGTCAAGAGTGCCACCTTGTCCTCCACGGAGAGCACGGCGTCT 2052

Qy    639 GluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArg 658
Db   2053 GGAATGCAGGATGAGGTCAGAGCGGTTGGCGAGGAGCAAAGGTCAGAGGAGCCAAAGCGC 2112

Qy    659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
Db   2113 AGGGTGGATACTTCAGTGTCTTGGGAGGCGTTGATTTGTGTCGGATCGTCCAAGAAGAGA 2172

Qy    679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
Db   2173 GCGAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGAACACTGGGAGGGGATGGC 2232

Qy    699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
Db   2233 CACAGAGCGGAGGAGGCTAGCAAAGACAAAGAAGCA------GATGCTCTTCCTGCCAGC 2286

Qy    719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
Db   2287 ACCCAGGAACAAGACCAAGCGCACGGAAGTTCCTCACCCGAGCCAGCTGGAAGCCCTTCT 2346

Qy    739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLys 758
Db   2347 GAAGGGGAGGGCGTCTCCACCTGGGAGTCATTTAAGAGATTAGTCACTCCACGAAAAAAA 2406

Qy    759 SerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGlu 778
Db   2407 TCCAAGTCAAAACTGGAAGAGAGAGCCGAAGACTCCGGTGCAGAGCAGTTGGCCTCCGAG 2466

Qy    779 IleGluProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArg 798
Db   2467 ATCGAACCAAGTAGAGAGGAATCTTGGGTTTCCATTAAGAAATTTATTCCTGGACGGCGG 2526

Qy    799 LysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGlu 818
Db   2527 AAGAAAAGGGCAGATGGGAAGCAAGAACAGGCCGCCGTTGAAGACTCGGGGCCAGGAGAG 2586

Qy    819 IleAsnGluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaVal 838
Db   2587 ATCAATGAGGACGACCCCGACGTCCCAGCTGTTGTGCCTCTGTCTGAGTACGATGCGGTA 2646

Qy    839 GluArgGluLysMetGlu---------AlaGlnGlyAsnThrGluLeuProGlnLeuLeu 855
Db   2647 GAGAGAGAGAAGCTGGAAGCGCAGCGAGCTCAGGAGAACGTGGAGCTGCCCCAGCTGAAG 2706

Qy    856 GlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrValSerValAla 875
Db   2707 GGGGCTGTGTATGTGTCTGAGGAGCTTAGTAAGACTCTGGTTCACACCGTGAGTGTCGCG 2766

Qy    876 ValIleAspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSer 895
Db   2767 GTCATTGATGGGACCAGGGCAGTCACCAGTGCCGAAGAGCGGTCCCCTTCGTGGATATCT 2826

Qy    896 AlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGluGlu 915
Db   2827 GCTTCCATGACAGAACCTCTTGAGCACGCAGAGGGAGTGGCCACACCGCCTGTTGGAGAG 2886

Qy    916 ValThrGluLysAspIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGlu 935
Db   2887 GTCACTGAAAAAGACATCACTGCAGAAGCAACTCCTGCACTCGCCCAGACTTTACCAGGG 2946

Qy    936 GlyLysAspAlaHisAspAspMetValThrSerGluValAspPheThrSerGluAlaVal 955
Db   2947 GGCAAAGATGCCCATGACGACATAGTCACCAGTGAAGTGGATTTTACCTCAGAAGCAGTG 3006
```

```
Qy    956 ThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAla 975
Db   3007 ACAGCCGCAGAAACCACAGAGGCGCTCCGCGCTGAAGAACTTACCGAAGCATCAGGGGCA 3066

Qy    976 GluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThr 995
Db   3067 GAAGAGACCACAGACATGGTGTCTGCAGTTTCCCAGCTGTCCGACTCCCCGGACACCACA 3126

Qy    996 GluGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGlu 1015
Db   3127 GAGGAAGCCACCCCAGTTCAGGAGGTAGAGGGTGGCATGCTAGATACGGAAGAACAGGAG 3186

Qy   1016 ArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnVal 1035
Db   3187 CGCCAGACGCAGGCCGTCCTCCAAGCCGTTGCAGACAAAGTGAAAGAGGACTCCCAGGTG 3246

Qy   1036 ProAlaThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluVal 1055
Db   3247 CCTGCAACCCAGACTCTGCAGAGAGCAGGGCCGAAAGCACTGGAGAAGGTGGAGGAGGTA 3306

Qy   1056 GluGluAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyPro 1075
Db   3307 GAGGAGGACTCCGAGGTGCTGGCTACCGAGAAAGAGAAGGATGTTGTGCCGGAAGGACCC 3366

Qy   1076 ValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrPro 1095
Db   3367 GTGCAGGAAGCTGAAACTGAGCATCTTGCACAGGGCTCCGAGACTGTACAGGCTACCCCA 3426

Qy   1096 GluSerLeuGluValProGluValThrAlaAspValAspHisValAlaThrCysGlnVal 1115
Db   3427 GAGAGCCTTGAAGTTCCTGAAGTCACAGAGGATGTAGACCGTGCCACCACATGCCAGGTT 3486

Qy   1116 IleLysLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThr 1135
Db   3487 ATCAAGCACCAGCAGCTGATGGAACAGGCTGTGGCCCCTGAGTCATCTGAAACCTTGACA 3546

Qy   1136 AspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGln 1155
Db   3547 GACAGTGAGACAAATGGAAGTACTCCCCTCGCAGATTCAGACACTCCAAACGGGACACAG 3606

Qy   1156 GlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGln 1175
Db   3607 CAAGACGAGACCGTTGACAGCCAGGACAGTAATGCCATTGCCGCCGTCAAGCAGTCACAG 3666

Qy   1176 ValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsn 1195
Db   3667 GTCACTGAAGAGGAGGCAGCTGCTGCTCAGACGGAGGGGCCTTCAACA---CCATCTAGT 3723

Qy   1196 ValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProThrGln 1215
Db   3724 TTTCCAGCCCAGGAAGAACACAGGGAAAAACCAGGAAGGGATGTTCTAGAACCCACA--- 3780

Qy   1216 GlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGly 1235
Db   3781 CAAGCGCTGGCTGCCGGGGCAGTGCCTATTCTGGCAAAGGCTGAGGTGGGTCAAGAGGGT 3840

Qy   1236 GluValAspTrpLeuAspGlyGluLysValLysGlu---------------GluGlnGlu 1250
Db   3841 GAGGCTGGCCAGTTTGATGGAGAAAAAGTCAAAGACGGACAGTGTGTTAAAGAACTGGAG 3900

Qy   1251 ValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGlu 1270
Db   3901 GTGCCTGTGCACACTGGACCCAACAGTCAAAAGACTGCTGACTTGACACGTGACAGTGAA 3960

Qy   1271 ValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSerLeuSerLeu 1290
Db   3961 GTAATGGAAGTGGCCAGATGTCAGGAAACTGAGAGTAATGAAGAACAGAGTATTAGCCCG 4020

Qy   1291 GluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGluGln 1310
Db   4021 GAGAAAAGAGAGATGGGAACCGACGTTGAAAAGGAGGAAACAGAGACCAAGACAGAGCAA 4080

Qy   1311 ValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLys 1330
Db   4081 GCCAGTGAAGAACATGAGCAGGAAACAGCTGCTCCTGAGCATGAAGGAACCCACCCTAAG 4140

Qy   1331 ProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGly 1350
Db   4141 CCAGTCCTGACAGCTGACATGCCTCACTCAGAGAGGGGAAAGGCACTGGGCAGCCTTGAA 4200

Qy   1351 GlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSer 1370
Db   4201 GGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGATTGCATAGAGGTTCAAGTTCAAAGC 4260

Qy   1371 LeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrValVal 1390
Db   4261 TCAGACACACCAGTCACTCAAACAACCGAAGCTGTGAAAAAGGTCGAAGAAACTGTGGCA 4320

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4321 ACTTCAGAGATGGATGAAAGTTTGGAGTGTGCAGGTGCGCAATCATTACCAGCTGAGAAG 4380

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4381 CTCTCCGAAACCGGTGGCTACGGGACTCTTCAGCATGGAGAGGACACCGTGCCCCAGGGG 4440

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4441 CCTGAGTCTCAGGCAGAGTCCATCCCAATAATAGTAACTCCTGCTCCTGAAAGCATCCTA 4500

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4501 CATTCTGACCTTCAAAGAGAAGTGAGCGCATCCCAGAAACAGAGATCAGATGAAGATAAC 4560

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4561 AAGCCAGATGCTGGTCCTGATGCTGCCGGCAAGGAGAGTGCAGCAAGAGAGAAAATCCTC 4620

Qy   1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
Db   4621 AGGGCTGAACCTGAGATCTTGGAACTTGAGAGTAAGAGCAATAAGATTGTCCAGAGTGTC 4680

Qy   1511 IleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaProGluThrHisAlaTyr 1530
Db   4681 ATCCAGACAGCCGTCGACCAGTTTGCACGTACAGAAACAGCCCCCGAAACCCACGCTTCT 4740

Qy   1531 AspSerGlnThrGlnValProAla-CysArgLeuAspSerArgGluProAsnArgCysTr 1550
Db   4741 GATTTACAGAATCAGGTTCCTGTGATGCAGGCTGACAGCCAGGGAGCACAACAGATGCTG 4800

Qy   1550 pThrLysMetLys---------------AspAlaLysMetLysHisProValProGlnPr 1565
Db   4801 GACAAAGATGAAAGCGACCTTCAAGTCTCCCCCCAAGATGGAACACTCAGTGCCGTAGCC 4860

Qy   1565 oArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLe 1585
Db   4861 CAGGAAGGACTTGCGGTTTCTGATAGTTCTGAAGGCATGAGCAAGGCTTCAGAAATGATC 4920

Qy   1585 uProArgLeuGlnLeuLysAlaProValSerLys 1596
Db   4921 ACCACGCTTGCAGTTGAAAGTGCCAGTGTCAAAG 4954
```

RESULT 8
US-08-978-277A-1
; Sequence 1, Application US/08978277A
; GENERAL INFORMATION:
; APPLICANT: Gelman, Irwin H.
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
; STREET: 30 Rockefeller Plaza
; CITY: New York
; STATE: NY
; COUNTRY: USA

```
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/978,277A
;      FILING DATE:
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5074 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-978-277A-1

Alignment Scores:
Pred. No.:               6.1e-300     Length:        5074
Score:                   6005.00      Matches:       1232
Percent Similarity:      97.33%       Conservative:  7
Best Local Similarity:   96.78%       Mismatches:    27
Query Match:             74.38%       Indels:        14
DB:                      13           Gaps:          4

US-09-902-432-4 (1-1596) x US-08-978-277A-1 (1-5074)

Qy     329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 61

Qy     349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 121

Qy     369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 181

Qy     389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 241

Qy     409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 301

Qy     429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 361

Qy     449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 421

Qy     469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 481

Qy     489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 541

Qy     509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 601

Qy     529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 661

Qy     549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 721

Qy     569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588
           ||||||||||||||||||   ||||||||||||||||||||||||||||||    |||||
Db     722 GGGCCGCTGGAAGCACCCAG-GATGGGGAAGCTGAGGAAGGAACTACTTCGT--GGAGAG 778

Qy     589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608
           |||||||||:::   |||||||||||||||||||||||||||||||||||||||||||||
Db     779 AAGAAGAGGAAGG--ATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 836

Qy     609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628
           |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     837 CGGTC-CGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 895

Qy     629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     896 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 955

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 1015

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATATAAGA 1075

Qy     689 GlyGlyProArgThrLeuGly-GlyAspSerHisArgAlaGluGluAlaSerLysAspLy 708
              |||||||||||||||||  |||||||||||||||||||||||||||||||||||||||
Db    1076 ---GGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAA 1132

Qy     708 sGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySe 728
           ||||||   |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1133 AGAAGCCGA-ACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAG 1191

Qy     728 rSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSe 748
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1192 TTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTC 1251

Qy     748 rPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla-- 767
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1252 ATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA-AG 1310

Qy     768 ----GluAspSerSerVal---GluGlnLeuSerThrGluIleGluProSerArgGluGl 785
               :::|||||||||     :::||||||||||||||||||||||    ||||||||
Db    1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy     785 uSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLy 805
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy     805 s-GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProA 825
           |  |||||||||||||||||||||||||||||||||||||||||||||||||:::|||:
Db    1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy     825 snValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluA 845
           ::||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
```

```
Db   1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy    845 laGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865
          ||||||||||    ||||||||          |||    |||   |||||||||||||||
Db   1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy    865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy    885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy    905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluG 925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy    925 luThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValT 945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy    945 hrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuA 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy    965 rgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaV 985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy    985 alSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValG 1005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   1005 luSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaV 1025
          ||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   1025 alAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrG 1045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   1045 lySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerG 1065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   1065 luLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuA 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   1085 laGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrA 1105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   1105 laAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnA 1125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   1125 laValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProL 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   1145 euAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspS 1165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   1165 erLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaG 1185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   1185 lnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluG 1205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   1205 luProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProV 1225
          |||||||||||||||||||||||||||||||||||||||     ||||||||||||||
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   1225 alLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysV 1245
          ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   1245 alLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspV 1265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   1265 alThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluV 1285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   1285 alGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgG 1305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   1305 luThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisG 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   1325 luGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysA 1345
          ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   1345 laLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleG 1365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   1365 luValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysV 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   1385 alIleGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisL 1405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   1405 euLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluA 1425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   1425 spThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProA 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   1445 laProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluA 1465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   1465 rgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrA 1485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   1485 laIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnL 1505
          |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   1505 ysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaP 1525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   1525 roGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgG 1545
          |||||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   1545 luProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnP 1565
          |||||||||||||||||||||||||||||||||    |||||||||||||||||||||||
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707
```

```
Qy    1565 roArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysL 1585
           ||||||||||||||||||||||||||||||||||||||||||||||     ||||||||||
Db    3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGC--TCAGCTCGGAAATGCT 3765

Qy    1585 euProArgLeuGlnLeuLysAlaProValSerLys 1596
           ||||||||||||||||||||||||||||||||||||
Db    3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 3800

RESULT 9
PCT-US97-06830-1
; Sequence 1, Application PC/TUS9706830
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  20
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US97/06830
;      FILING DATE:
;      CLASSIFICATION:
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  08/665,401
;      FILING DATE:  18-JUN-1996
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/34008
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5134 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
PCT-US97-06830-1

Alignment Scores:
Pred. No.:                6.18e-300     Length:         5134
Score:                    6005.00       Matches:        1232
Percent Similarity:       97.33%        Conservative:   7
Best Local Similarity:    96.78%        Mismatches:     27
Query Match:              74.38%        Indels:         14
DB:                       1             Gaps:           4

US-09-902-432-4 (1-1596) x PCT-US97-06830-1 (1-5134)

Qy     329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 61

Qy     349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 121

Qy     369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 181

Qy     389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 241

Qy     409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 301

Qy     429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 361

Qy     449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 421

Qy     469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 481

Qy     489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 541

Qy     509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 601

Qy     529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 661

Qy     549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 721

Qy     569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588
           |||||||||||||||||| |||||||||||||||||||||||||||||||   ||||||
Db     722 GGGCCGCTGGAAGCACCCAG-GATGGGGAAGCTGAGGAAGGAACTACTTCGT--GGAGAG 778

Qy     589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608
           ||||||||:::  |||||||||||||||||||||||||||||||||||||||||||||||
Db     779 AAGAAGAGGAAGG--ATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 836

Qy     609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628
           |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     837 CGGTC-CGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 895

Qy     629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     896 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 955

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 1015

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATATAAGA 1075

Qy     689 GlyGlyProArgThrLeuGly-GlyAspSerHisArgAlaGluGluAlaSerLysAspLy 708
              ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db    1076 ---GGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAA 1132

Qy     708 sGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySe 728
           ||||||     |||||||||||||||||||||||||||||||||||||||||||||||||
Db    1133 AGAAGCCGA-ACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAG 1191

Qy     728 rSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSe 748
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1192 TTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTC 1251

Qy      748 rPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla-- 767
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1252 ATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA-AG 1310

Qy      768 ----GluAspSerSerVal---GluGlnLeuSerThrGluIleGluProSerArgGluGl 785
                :::||||||||      :::||||||||||||||||||||   ||||||||
Db     1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy      785 uSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLy 805
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy      805 s-GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProA 825
            | ||||||||||||||||||||||||||||||||||||||||||||||||||:::|||:
Db     1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy      825 snValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluA 845
            ::||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
Db     1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy      845 laGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865
            ||||||||||  ||||||||||          |||   |||   ||||||||||||||||
Db     1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy      865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy      885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy      905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluG 925
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy      925 luThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValT 945
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy      945 hrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuA 965
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy      965 rgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaV 985
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy      985 alSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValG 1005
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy     1005 luSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaV 1025
            ||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy     1025 alAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrG 1045
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy     1045 lySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerG 1065
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy     1065 luLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuA 1085
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy     1085 laGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrA 1105
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy     1105 laAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnA 1125
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy     1125 laValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProL 1145
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy     1145 euAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspS 1165
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy     1165 erLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaG 1185
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy     1185 lnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluG 1205
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy     1205 luProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProV 1225
            ||||||||||||||||||||||||||||||||||||||||||||   |||||||||||||
Db     2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy     1225 alLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysV 1245
            ||         |||||||||||||||||||||||||||||||||||||||||||||||||
Db     2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy     1245 alLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspV 1265
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy     1265 alThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluV 1285
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy     1285 alGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgG 1305
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy     1305 luThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisG 1325
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy     1325 luGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysA 1345
            ||        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy     1345 laLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleG 1365
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy     1365 luValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysV 1385
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy     1385 alIleGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisL 1405
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy     1405 euLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluA 1425
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy     1425 spThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProA 1445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy     1445 laProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluA 1465
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407
```

```
Qy    1465 rgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrA 1485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy    1485 laIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnL 1505
           |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy    1505 ysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaP 1525
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy    1525 roGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgG 1545
           ||||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db    3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy    1545 luProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnP 1565
           ||||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db    3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy    1565 roArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysL 1585
           ||||||||||||||||||||||||||||||||||||||||||||        ||||||||||
Db    3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGC--TCAGCTCGGAAATGCT 3765

Qy    1585 euProArgLeuGlnLeuLysAlaProValSerLys 1596
           |||||||||||||||||||||||||||||||||||
Db    3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 3800
```

RESULT 10
US-08-665-401-1
; Sequence 1, Application US/08665401
; GENERAL INFORMATION:
;   APPLICANT: Gelman, Irwin H.
;   TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;   NUMBER OF SEQUENCES: 20
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Brumbaugh, Graves, Donohue & Raymond
;   STREET: 30 Rockefeller Plaza
;   CITY: New York
;   STATE: NY
;   COUNTRY: USA
;   ZIP: 10112-0228
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Diskette
;   COMPUTER: IBM Compatible
;   OPERATING SYSTEM: DOS
;   SOFTWARE: FastSEQ Version 1.5
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/665,401
;   FILING DATE: 18-JUN-1996
;   CLASSIFICATION: 514
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Clark, Richard S
;   REGISTRATION NUMBER: 26,154
;   REFERENCE/DOCKET NUMBER: A30558 - 165/34008
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 212-408-2558
;   TELEFAX: 212-765-2519
;   TELEX:
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 5134 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   HYPOTHETICAL: NO
;   ANTI-SENSE: NO
;   FRAGMENT TYPE:
;   ORIGINAL SOURCE:
US-08-665-401-1

```
Alignment Scores:
Pred. No.:              6.18e-300     Length:        5134
Score:                  6005.00       Matches:       1232
Percent Similarity:     97.33%        Conservative:  7
Best Local Similarity:  96.78%        Mismatches:    27
Query Match:            74.38%        Indels:        14
DB:                     10            Gaps:          4
```

US-09-902-432-4 (1-1596) x US-08-665-401-1 (1-5134)

```
Qy     329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 61

Qy     349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 121

Qy     369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 181

Qy     389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 241

Qy     409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 301

Qy     429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 361

Qy     449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 421

Qy     469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 481

Qy     489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 541

Qy     509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 601

Qy     529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 661

Qy     549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 721

Qy     569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588
           ||||||||||||||||   |||||||||||||||||||||||||||||    ||||||
Db     722 GGGCCGCTGGAAGCACCCAG-GATGGGGAAGCTGAGGAAGGAACTACTTCGT--GGAGAG 778

Qy     589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608
           |||||||||:::   ||||||||||||||||||||||||||||||||||||||||||||
Db     779 AAGAAGAGGAAGG--ATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 836

Qy     609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628
           |||   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     837 CGGTC-CGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 895
```

```
Qy     629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     896 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 955

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 1015

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATATAAGA 1075

Qy     689 GlyGlyProArgThrLeuGly-GlyAspSerHisArgAlaGluGluAlaSerLysAspLy 708
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    1076 ---GGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAA 1132

Qy     708 sGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySe 728
           ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1133 AGAAGCCGA-ACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAG 1191

Qy     728 rSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSe 748
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1192 TTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTC 1251

Qy     748 rPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla-- 767
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1252 ATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA-AG 1310

Qy     768 ----GluAspSerSerVal---GluGlnLeuSerThrGluIleGluProSerArgGluGl 785
               :::||||||||         :::||||||||||||||||||||    |||||||
Db    1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy     785 uSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLy 805
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy     805 s-GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProA 825
           | |||||||||||||||||||||||||||||||||||||||||||||||||:::|||:
Db    1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy     825 snValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluA 845
           ::|||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db    1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy     845 laGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865
           ||||||||||    |||||||||          |||    |||   |||||||||||||||
Db    1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy     865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy     885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy     905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluG 925
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy     925 luThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValT 945
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy     945 hrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuA 965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy     965 rgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaV 985
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy     985 alSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValG 1005
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy    1005 luSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaV 1025
           ||   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy    1025 alAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrG 1045
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy    1045 lySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerG 1065
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy    1065 luLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuA 1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy    1085 laGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrA 1105
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy    1105 laAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnA 1125
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy    1125 laValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProL 1145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy    1145 euAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspS 1165
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy    1165 erLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaG 1185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy    1185 lnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluG 1205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy    1205 luProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProV 1225
           ||||||||||||||||||||||||||||||||||||||    ||||||||||||||||||
Db    2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy    1225 alLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysV 1245
           ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy    1245 alLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspV 1265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy    1265 alThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluV 1285
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy    1285 alGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgG 1305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy    1305 luThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisG 1325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy    1325 luGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysA 1345
           ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy    1345 laLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleG 1365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   1365 luValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysV 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   1385 alIleGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisL 1405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   1405 euLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluA 1425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   1425 spThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProA 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   1445 laProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluA 1465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   1465 rgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrA 1485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   1485 laIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnL 1505
          ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   1505 ysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaP 1525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   1525 roGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgG 1545
          ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   1545 luProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnP 1565
          ||||||||||||||||||||||||||||    ||||||||||||||||||||||||||||
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   1565 roArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysL 1585
          ||||||||||||||||||||||||||||||||||||||||||          ||||||||
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGC--TCAGCTCGGAAATGCT 3765

Qy   1585 euProArgLeuGlnLeuLysAlaProValSerLys 1596
          |||||||||||||||||||||||||||||||||||
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 3800

RESULT 11
US-09-902-432-1
; Sequence 1, Application US/09902432
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 5134
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-902-432-1

  Alignment Scores:
  Pred. No.:               6.18e-300      Length:        5134
  Score:                   6005.00        Matches:       1232
  Percent Similarity:      97.33%         Conservative:  7
  Best Local Similarity:   96.78%         Mismatches:    27
  Query Match:             74.38%         Indels:        14
  DB:                      34             Gaps:          4

US-09-902-432-4 (1-1596) x US-09-902-432-1 (1-5134)

Qy    329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 61

Qy    349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     62 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 121

Qy    369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    122 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 181

Qy    389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    182 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 241

Qy    409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    242 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 301

Qy    429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    302 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 361

Qy    449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    362 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 421

Qy    469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    422 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 481

Qy    489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    482 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 541

Qy    509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    542 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 601

Qy    529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    602 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 661

Qy    549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    662 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 721

Qy    569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588
          |||||||||||||||||   |||||||||||||||||||||||||||||    |||||||
Db    722 GGGCCGCTGGAAGCACCCAG-GATGGGGAAGCTGAGGAAGGAACTACTTCGT--GGAGAG 778

Qy    589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608
          ||||||||||:::     ||||||||||||||||||||||||||||||||||||||||||
Db    779 AAGAAGAGGAAGG--ATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 836

Qy    609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628
          |||    |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    837 CGGTC-CGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 895

Qy    629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     896 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 955

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 1015

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATATAAGA 1075

Qy     689 GlyGlyProArgThrLeuGly-GlyAspSerHisArgAlaGluGluAlaSerLysAspLy 708
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db    1076 ---GGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAA 1132

Qy     708 sGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySe 728
           ||||||    ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1133 AGAAGCCGA-ACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAG 1191

Qy     728 rSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSe 748
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1192 TTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTC 1251

Qy     748 rPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla-- 767
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1252 ATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA-AG 1310

Qy     768 ----GluAspSerSerVal---GluGlnLeuSerThrGluIleGluProSerArgGluGl 785
                :::||||||||       :::|||||||||||||||||||     |||||||
Db    1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy     785 uSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLy 805
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy     805 s-GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProA 825
           | |||||||||||||||||||||||||||||||||||||||||||||||||:::|||:
Db    1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy     825 snValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluA 845
           ::|||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db    1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy     845 laGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865
           ||||||||||   ||||||||          |||    |||   |||||||||||||||
Db    1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy     865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy     885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy     905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluG 925
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy     925 luThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValT 945
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy     945 hrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuA 965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy     965 rgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaV 985
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy     985 alSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValG 1005
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy    1005 luSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaV 1025
           ||   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy    1025 alAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrG 1045
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy    1045 lySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerG 1065
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy    1065 luLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuA 1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy    1085 laGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrA 1105
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy    1105 laAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnA 1125
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy    1125 laValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProL 1145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy    1145 euAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspS 1165
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy    1165 erLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluAlaAlaThrAlaG 1185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy    1185 lnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluG 1205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy    1205 luProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProV 1225
           ||||||||||||||||||||||||||||||||||||||     ||||||||||||||||
Db    2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy    1225 alLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysV 1245
           ||       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy    1245 alLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspV 1265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy    1265 alThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluV 1285
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy    1285 alGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgG 1305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy    1305 luThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisG 1325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy    1325 luGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysA 1345
           ||   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy    1345 laLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleG 1365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107
```

```
Qy   1365 luValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysV 1385
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   1385 alIleGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisL 1405
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   1405 euLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluA 1425
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   1425 spThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProA 1445
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   1445 laProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluA 1465
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   1465 rgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrA 1485
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   1485 laIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnL 1505
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   1505 ysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaP 1525
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   1525 roGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgG 1545
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   1545 luProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnP 1565
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   1565 roArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysL 1585
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGC--TCAGCTCGGAAATGCT 3765

Qy   1585 euProArgLeuGlnLeuLysAlaProValSerLys 1596
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 3800

RESULT 12
US-09-770-174-4018
; Sequence 4018, Application US/09770174
; GENERAL INFORMATION:
;  APPLICANT: Shyjan, Andrew W.
;  TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
;  TITLE OF INVENTION:  THEREFOR
;  FILE REFERENCE: 1600.2057-001
;  CURRENT APPLICATION NUMBER: US/09/770,174
;  CURRENT FILING DATE:  2001-01-26
;  PRIOR APPLICATION NUMBER: US 60/178,835
;  PRIOR FILING DATE: 2000-01-28
;  NUMBER OF SEQ ID NOS: 4376
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4018
;   LENGTH: 7651
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-770-174-4018

Alignment Scores:
Pred. No.:               1.47e-213      Length:         7651
Score:                   4334.50        Matches:        969
Percent Similarity:      69.07%         Conservative:   190
Best Local Similarity:   57.75%         Mismatches:     409

Query Match:             53.69%         Indels:         110
DB:                      30             Gaps:           33

US-09-902-432-4 (1-1596) x US-09-770-174-4018 (1-7651)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    183 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 242

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    243 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 302

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    303 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 362

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    363 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 422

Qy     77 GlnGluGlyGln-----------------------------------------GluGlu 82
Db    423 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 482

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    483 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 542

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    543 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 602

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    603 GAGACACCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 662

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    663 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 722

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    723 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 782

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    783 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 833

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    834 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 887

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    888 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 947

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    948 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1004

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1005 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1064

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1065 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1124

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1125 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1184
```

```
Qy    323 ValAspGluGluLysGluLysThrGluProAlaSerGluGlu-------------- 337
Db   1185 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1244

Qy    338 -----------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1245 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1304

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1305 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1364

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1365 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1424

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1425 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1475

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1476 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1520

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1521 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1580

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1581 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1640

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1641 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1700

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1701 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1760

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1761 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1817

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1818 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1877

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1878 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1937

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1938 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 1997

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   1998 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2057

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2058 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2117

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2118 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2177

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2178 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2237

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2238 GGATCATCCAAGAAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2297

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2298 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2357

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2358 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2417

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2418 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2477

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2478 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2537

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2538 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2597

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2598 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGAAACAAGAACAA 2657

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2658 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2717

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2718 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2777

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2778 CAAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2837

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2838 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2897

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2898 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2957

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2958 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3017

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3018 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3077

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3078 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3137

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3138 GGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3197

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3198 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3257

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3258 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3317

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr----------------- 1038
```

```
Db   3318 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3377

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
             |||   ||||||||                       :::   |||     |||||
Db   3378 CTTCAGCCTGTGCAGAGAGCA-----------------GAGGCAGAAAGACCAGAAGAG 3419

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
             :::|||        :::||||||  ||||||:::   |||              |||
Db   3420 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3479

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
          ||||||       |||         ||||||           ||||||  ||||||||||||
Db   3480 GAGGCAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3539

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
             |||      |||:::||||||       ::::::            |||   ||||||
Db   3540 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3599

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
          |||                     :::|||  |||::::::    ||||||||||||:::
Db   3600 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3659

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
          |||:::|||    ||||||   |||||||||||||||:::||||||||||||:::||||||
Db   3660 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3719

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
             |||             ||||||:::||||||    ::::::      :::
Db   3720 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3779

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
             |||:::        ||||||      ||||||  ||||||      ||||||||||||
Db   3780 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3839

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
             |||       |||:::::           |||||||||       |||:::
Db   3840 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3896

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
                |||   ||||||   |||   :::|||::::::          |||   :::|||
Db   3897 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3956

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
          :::|||||||||         |||      |||   |||         ||||||   |||
Db   3957 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4010

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
                         |||      |||      :::   :::|||         :::
Db   4011 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4070

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
          :::|||      :::|||               |||         |||   |||::::::
Db   4071 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4130

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
             ::::::      |||:::||||||               |||      |||   |||
Db   4131 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4190

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
          |||         ||||||:::||||||   |||   :::|||   .   |||:::||||||
Db   4191 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4250

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
                |||   |||||||||               |||         :::   :::
Db   4251 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4310

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
          |||:::::::::|||         :::      |||   :::   ||||||   ||||||
Db   4311 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4370

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
          |||         |||::::::|||   |||      ::::::||||||||||||   :::
Db   4371 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4430
```

```
Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
             :::   |||   ||||||   |||   |||:::   :::        ||||||
Db   4431 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4490

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
             |||   ||||||||||||:::   |||      ||||||||||||:::      ||||||
Db   4491 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4550

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
          ||||||            :::               ||||||   ||||||   |||
Db   4551 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4610

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
          |||:::   ||||||:::|||   |||:::||||||:::         :::      |||
Db   4611 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4670

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
             ||||||:::||||||   :::   |||   :::   :::|||:::|||      |||
Db   4671 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4730

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
          :::            :::         |||   |||   |||||||||:::   |||
Db   4731 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTA 4784

Qy   1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
          :::   |||      ||||||||||||:::||||||:::|||:::|||   |||:::
Db   4785 GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATC 4844

Qy   1511 IleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThrHisAla 1529
          |||||||||||||||||||||   ||||||      |||||||||   |||
Db   4845 ATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACG 4904

Qy   1530 TyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCys 1549
             :::   |||||||||         :::   ||||||::::::      :::
Db   4905 TCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAA 4964

Qy   1550 TrpThrLys--------------MetLysAspAla-LysMetLysHisProValProGl 1564
                              |||   |||   :::||||||||   :::   ||
Db   4965 ACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCA 5024

Qy   1564 nProArgGlu--------------------AspLeuGlnValLeuThrValLeuGluAl 1577
           ||||:::                       |||   :::|||         :::
Db   5025 GCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGAC 5084

Qy   1577 aTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
               :::|||:::|||            |||:::   |||   |||:::
Db   5085 ATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5136
```

RESULT 13
US-60-278-232-10143
; Sequence 10143, Application US/60278232
; GENERAL INFORMATION:
; APPLICANT: Morris, MacDonald
; APPLICANT: Lal, Preeti
; APPLICANT: Diep, Dinh
; TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
; TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide
; TITLE OF INVENTION: Polymorphisms Identified Thereby
; FILE REFERENCE: GX-0011 P
; CURRENT APPLICATION NUMBER: US/60/278,232
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 12,557
; SOFTWARE: PERL Program
; SEQ ID NO 10143
;   LENGTH: 6603
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 445786.12
US-60-278-232-10143

```
Alignment Scores:
Pred. No.:                1.43e-213      Length:          6603
Score:                    4333.50        Matches:         968
Percent Similarity:       69.07%         Conservative:    191
Best Local Similarity:    57.69%         Mismatches:      409
Query Match:              53.68%         Indels:          110
DB:                       71             Gaps:            33

US-09-902-432-4 (1-1596) x US-60-278-232-10143 (1-6603)

Qy       1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db     193 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 252

Qy      20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db     253 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 312

Qy      39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db     313 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 372

Qy      57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db     373 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 432

Qy      77 GlnGluGlyGln------------------------------------------GluGlu 82
Db     433 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 492

Qy      83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db     493 GAAGTCATTGTCACAGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCAAAAGAGACTCC 552

Qy     103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db     553 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 612

Qy     123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db     613 GAGACACCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 672

Qy     143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db     673 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 732

Qy     163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db     733 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 792

Qy     183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db     793 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 843

Qy     203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db     844 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 897

Qy     223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db     898 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 957

Qy     243 AlaGluSerAspGlnAlaAlaGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db     958 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1014

Qy     263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db    1015 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1074

Qy     283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db    1075 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1134

Qy     303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db    1135 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1194

Qy     323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db    1195 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1254

Qy     338 -----------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db    1255 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1314

Qy     354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db    1315 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1374

Qy     374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db    1375 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1434

Qy     393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db    1435 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1485

Qy     413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db    1486 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1530

Qy     433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db    1531 TTGGTTGAAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1590

Qy     453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db    1591 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1650

Qy     473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db    1651 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1710

Qy     493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db    1711 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1770

Qy     513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db    1771 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1827

Qy     533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db    1828 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1887

Qy     553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db    1888 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1947

Qy     573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db    1948 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2007

Qy     593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db    2008 GGTGTCACTCCCTGGGCTTCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2067

Qy     613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db    2068 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2127

Qy     633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db    2128 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2187
```

```
Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2188 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2247

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2248 GGATCATCCAAGAAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2307

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2308 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2367

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2368 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGCAGGGAAGTTCCTCCCCGGAG 2427

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2428 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2487

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2488 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2547

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2548 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2607

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2608 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2667

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2668 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2727

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2728 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2787

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2788 CAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2847

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2848 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2907

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2908 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2967

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2968 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3027

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3028 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3087

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3088 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3147

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3148 GGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3207

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3208 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3267

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3268 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGACA 3327

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3328 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3387

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3388 CTTCAGCCTGTGCAGAGAGCA-----------------GAGGCAGAAAGACCAGAAGAG 3429

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3430 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3489

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3490 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3549

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3550 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3609

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3610 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3669

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3670 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3729

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3730 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3789

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3790 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3849

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3850 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3906

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3907 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3966

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3967 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4020

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4021 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4080

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4081 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4140

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4141 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4200

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4201 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4260

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4261 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4320

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
```

```
Db   4321 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4380

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4381 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4440

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4441 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4500

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4501 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4560

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4561 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4620

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4621 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4680

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4681 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4740

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4741 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTA 4794

Qy   1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
Db   4795 GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATC 4854

Qy   1511 IleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThrHisAla 1529
Db   4855 ATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACG 4914

Qy   1530 TyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCys 1549
Db   4915 TCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAA 4974

Qy   1550 TrpThrLys---------------MetLysAspAla-LysMetLysHisProValProGl 1564
Db   4975 ACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCA 5034

Qy   1564 nProArgGlu---------------------AspLeuGlnValLeuThrValLeuGluAl 1577
Db   5035 GCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGAC 5094

Qy   1577 aTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   5095 ATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5146
```

RESULT 14
US-60-245-228-517/c
; Sequence 517, Application US/60245228
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN CYCLASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN CYCLASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL000878
; CURRENT APPLICATION NUMBER: US/60/245,228
; CURRENT FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 630
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 517
; LENGTH: 6607
; TYPE: DNA
; ORGANISM: HUMAN
US-60-245-228-517

```
Alignment Scores:
Pred. No.:                 3.7e-213      Length:         6607
Score:                     4325.50       Matches:        969
Percent Similarity:        68.83%        Conservative:   188
Best Local Similarity:     57.64%        Mismatches:     408
Query Match:               53.58%        Indels:         116
DB:                        68            Gaps:           33

US-09-902-432-4 (1-1596) x US-60-245-228-517 (1-6607)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db   6413 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 6354

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAla 39
Db   6353 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTGGGCCGAG---------GCGGCG 6303

Qy     40 GlyAspProAlaAspAla------------------AspProAlaThrLysLeuProGln 53
Db   6302 CCAGACCCCCCCGCGGCCCCAACAATCGCTGCCTGGGACCCCGCCACCAAGCTCCTACAG 6243

Qy     54 LysAsnGlyGlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGln 73
Db   6242 AAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAG 6183

Qy     74 GluGluAsnGlnGluGlyGln--------------------------------------- 80
Db   6182 GAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAG 6123

Qy     81 ---GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGlu 99
Db   6122 GAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAA 6063

Qy    100 LysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAsp 119
Db   6062 AGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGAT 6003

Qy    120 GlyGlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGlu 139
Db   6002 GGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAA 5943

Qy    140 GluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLys 159
Db   5942 GAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAG 5883

Qy    160 PheValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGln 179
Db   5882 TTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAG 5823

Qy    180 LeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAsp 199
Db   5822 CTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGAC 5772

Qy    200 HisGlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeu 219
Db   5771 CACCAGGACCCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCC 5718

Qy    220 LysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIle 239
Db   5717 AAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATT 5658

Qy    240 ProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGlu 259
Db   5657 TCTCCCCCAGCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAG 5601

Qy    260 LysGlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGlu 279
Db   5600 AAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAA 5541

Qy    280 ThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThr 299
Db   5540 ACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACC 5481
```

```
Qy    300 SerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGlu 319
Db   5480 AGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAG 5421

Qy    320 AlaGluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu------ 337
Db   5420 CCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACC 5361

Qy    338 --------------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSer 350
Db   5360 GCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCA 5301

Qy    351 AlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSer 370
Db   5300 GCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCT 5241

Qy    371 SerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHis 390
Db   5240 TCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCAC 5181

Qy    391 Gln---GluValValAlaGluValHisValSerThrValGluLysThrGluGluGluGln 409
Db   5180 CAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG 5121

Qy    410 GlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuPro 429
Db   5120 -----------------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCA 5085

Qy    430 ProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeu 449
Db   5084 GCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTG 5025

Qy    450 MetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSer 469
Db   5024 GTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGT 4965

Qy    470 ProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeu 489
Db   4964 CCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTG 4905

Qy    490 SerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSer 509
Db   4904 TCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACT 4845

Qy    510 GlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGlu 529
Db   4844 GGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAG 4788

Qy    530 GluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLys 549
Db   4787 GAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAG 4728

Qy    550 GlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGly 569
Db   4727 GGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGC 4668

Qy    570 ProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLys 589
Db   4667 TTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAA 4608

Qy    590 LysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArg 609
Db   4607 AAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGT 4548

Qy    610 ValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThr 629
Db   4547 GTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACC 4488

Qy    630 LeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGlu 649
Db   4487 TTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGAGCGTGGAA 4428

Qy    650 GluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeu 669
Db   4427 GAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTA 4368

Qy    670 IleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGly 689
Db   4367 ATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGG 4308

Qy    690 GlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGlu 709
Db   4307 GGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAG 4248

Qy    710 AlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSer 729
Db   4247 ACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCC 4188

Qy    730 SerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPhe 749
Db   4187 TCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTT 4128

Qy    750 LysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp 769
Db   4127 AAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGAC 4068

Qy    770 -----------SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGlu 785
Db   4067 TCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAA 4008

Qy    786 SerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLys 805
Db   4007 TCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAA 3948

Qy    806 GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsn 825
Db   3947 CAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGAT 3888

Qy    826 ValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu--- 844
Db   3887 GTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCA 3828

Qy    845 ------AlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGlu 862
Db   3827 CAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAG 3768

Qy    863 GluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAla 882
Db   3767 GAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCA 3708

Qy    883 ValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeu 902
Db   3707 GCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGGACCTCTT 3648

Qy    903 GluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIle 922
Db   3647 GAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATT 3588

Qy    923 Ala---GluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAsp 941
Db   3587 GCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGC 3528

Qy    942 AspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSer 961
Db   3527 GACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCA 3468

Qy    962 GluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMet 981
Db   3467 GGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATG 3408

Qy    982 ValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProVal 1001
Db   3407 GTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTG 3348

Qy   1002 GlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIle 1021
```

```
Db   3347 CAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTC 3288

Qy   1022 LeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThr--------- 1038
Db   3287 CTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCA 3228

Qy   1039 ------------GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGlu 1054
Db   3227 GAAGATGTGCTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGA 3186

Qy   1055 ValGluGluAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGly 1074
Db   3185 CCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTA 3126

Qy   1075 ProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThr 1094
Db   3125 GATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACC 3066

Qy   1095 ProGluSerLeuGlu---ValProGluValThrAlaAspValAspHis---------Val 1110
Db   3065 CCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTA 3006

Qy   1111 AlaThrCysGlnVal---------------IleLysLeuGlnGlnLeu---MetGluGln 1124
Db   3005 ACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAG 2946

Qy   1125 AlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrPro 1144
Db   2945 GCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCC 2886

Qy   1145 LeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAsp 1164
Db   2885 GTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAG 2826

Qy   1165 SerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAla 1184
Db   2825 GAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCA 2766

Qy   1185 GlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGlu 1204
Db   2765 CAGAAAGAGAGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 2709

Qy   1205 GluProGly---ArgAspValLeuGluProThrGlnGluLeuThrAlaAlaAlaVal 1223
Db   2708 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 2649

Qy   1224 ProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGlu 1243
Db   2648 TCCATTCTGTCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAG 2595

Qy   1244 LysValLys---------------GluGluGlnGluValPheValHisSerGly------ 1256
Db   2594 AAAACCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACA 2535

Qy   1257 ProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGly 1276
Db   2534 GTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAA 2475

Qy   1277 CysGlnGluLysGluSerThrGluValGlnSer---------------LeuSerLeuGlu 1291
Db   2474 TGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTG 2415

Qy   1292 GluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnVal 1311
Db   2414 GAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTG 2355

Qy   1312 SerGluGluGly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLys 1330
Db   2354 AATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAG 2295

Qy   1331 ProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGly 1350
Db   2294 CTCCTCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAA 2235

Qy   1351 GlySer---ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGln 1369
Db   2234 GGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAG 2175

Qy   1370 SerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIle------Glu 1387
Db   2174 AGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAA 2115

Qy   1388 ThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuPro 1407
Db   2114 ACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTG 2055

Qy   1408 AlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrVal 1427
Db   2054 GAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTG 1995

Qy   1428 ProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGlu 1447
Db   1994 CCCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAG 1935

Qy   1448 SerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGlu 1467
Db   1934 AAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGAT 1875

Qy   1468 GluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGlu 1487
Db   1874 GAAGTCGATGAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAG 1821

Qy   1488 LysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleVal 1507
Db   1820 GAGGATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTC 1761

Qy   1508 LeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGlu 1526
Db   1760 CAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAA 1701

Qy   1527 ThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluPro 1546
Db   1700 ATGTTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCT 1641

Qy   1547 AsnArgCysTrpThrLys--------------MetLysAspAla-LysMetLysHisPr 1561
Db   1640 GGACAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCA 1581

Qy   1561 oValProGlnProArgGlu--------------------AspLeuGlnValLeuThrVa 1574
Db   1580 ATTACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATT 1521

Qy   1574 lLeuGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVa 1594
Db   1520 TCCAAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACT 1461

Qy   1594 l 1594
Db   1460 G 1460
```

RESULT 15
US-09-644-867-6364
; Sequence 6364, Application US/09644867
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; APPLICANT: Donovan, Michael J.
; APPLICANT: Leiby, Kevin R.
; APPLICANT: Culpepper, Janice A.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID MOLECULES AND USES
; TITLE OF INVENTION: THEREFOR
; FILE REFERENCE: 1600.1173-001
; CURRENT APPLICATION NUMBER: US/09/644,867
; CURRENT FILING DATE: 2000-08-28
; PRIOR APPLICATION NUMBER: 60/151,061
; PRIOR FILING DATE: 1999-08-27

```
;  NUMBER OF SEQ ID NOS:  8090
;  SOFTWARE:  FastSEQ for Windows Version 4.0
; SEQ ID NO 6364
;   LENGTH: 8660
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)...(8660)
;   OTHER INFORMATION: n = A,T,C or G
US-09-644-867-6364

Alignment Scores:
Pred. No.:                1.5e-212      Length:         8660
Score:                    4316.00       Matches:        967
Percent Similarity:       68.89%        Conservative:   189
Best Local Similarity:    57.63%        Mismatches:     411
Query Match:              53.46%        Indels:         111
DB:                       25            Gaps:           34

US-09-902-432-4 (1-1596) x US-09-644-867-6364 (1-8660)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
          ||||||||||||||||||||||||||||||||||||||||||   |||   ||||||
Db    183 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 242

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
          ||||||:::|||          ||||||  ||||||:::||||||||:::
Db    243 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 302

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
                ||||||||  |||      ||||||||||||||||  ||||||||||||
Db    303 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 362

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
          ||||||||::::::||||||||||||||||||      ::::::   :::||||||    :::
Db    363 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 422

Qy     77 GlnGluGlyGln------------------------------------------GluGlu 82
                ||||||                                          ||||||
Db    423 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 482

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
          ||||||:::         :::||||||||||||:::||||||||||||    |||:::|||
Db    483 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 542

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
             :::|||||||||         |||      |||    ||||||||    ||||||||||||
Db    543 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 602

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
          |||             ||||||||||||||||:::||||||:::|||:::||||||:::
Db    603 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 659

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
          ||||||    ||||||||||||||||||:::||||||||||||||||||||||||||||||
Db    660 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 719

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
          ||||||||||||||||||||||||||||||    ||||||   ||||||||||||||||||||
Db    720 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 779

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
          |||||||||||||||||||||||||||           ||||||||||||||||||:::
Db    780 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 830

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
          ||||||:::           ||||||:::||||||||||||||||||    |||||||||
Db    831 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 884

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
          |||||||||   |||    |||||||||:::|||||||||         ||||||
Db    885 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 944

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
          |||||||||      ||||||      ||||||   |||:::||||||||||||||||||||
Db    945 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1001

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
          |||||||||:::||||||   |||||||||:::|||||||||   |||||||||   |||
Db   1002 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1061

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
          :::||||||||||||||||||   ||||||||||||||||||||||||||||||||:::
Db   1062 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1121

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
          |||   |||||||||::::::|||   :::||||||:::||||||||||||   ||||||
Db   1122 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1181

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
          ||||||   ||||||      |||   |||   ||||||||||:::
Db   1182 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1241

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
                      |||||||||||||||      :::   |||||||||||:::|||
Db   1242 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1301

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
          ||||||||||||||||||   |||:::||||||         :::      ||||||||
Db   1302 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1361

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
          |||   |||||||||||||||||||||||||||||||||:::|||   ||||||   |||
Db   1362 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1421

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
          |||||||||||||||||||||||||||||||||:::      |||||||||
Db   1422 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAACCGAAGAGCAG--------- 1472

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
                         ||||||   |||      |||:::|||   |||:::
Db   1473 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1517

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
          |||   |||         ||||||||||||||||||||:::||||||:::|||
Db   1518 TTGGTTGAAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1577

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
          :::|||   ||||||||||||   |||   ||||||      ||||||||||:::|||
Db   1578 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1637

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
          |||   |||   |||   |||||||||:::||||||||||||||||||||||||||||||
Db   1638 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1697

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
          ||||||:::||||||||||||||||||||||||||||||:::|||:::|||||||||
Db   1698 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1757

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
          ||||||||||||||||||||||||||||||   ||||||||||||||||   |||
Db   1758 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1814

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
          |||:::      :::      :::||||||:::|||   :::||||||||||||||||||
Db   1815 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1874

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
          ||||||||||||||||||||||||||||||   ||||||||||||||||||   |||
Db   1875 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1934

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
                ||||||||||||||||||||||||   ||||||||||||||||||||||||||||
Db   1935 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 1994
```

```
Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   1995 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2054

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2055 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2114

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2115 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2174

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2175 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2234

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2235 GGATCATCCAAGAAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2294

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2295 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2354

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2355 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2414

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2415 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2474

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2475 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2534

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2535 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2594

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2595 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGAAACAAGAACAA 2654

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2655 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2714

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2715 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2774

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2775 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2834

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2835 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2894

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2895 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2954

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2955 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3014

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3015 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACACAGAGAGGCCCGGGGCGACACGGTC 3074

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3075 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3134

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3135 GGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3194

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3195 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3254

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3255 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3314

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3315 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3374

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3375 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3416

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3417 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3476

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3477 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3536

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3537 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3596

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3597 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3656

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3657 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3716

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3717 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3776

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3777 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3836

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3837 AGGCCTCCAGCA---CCTTCCAGTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3893

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3894 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3953

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3954 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4007

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4008 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4067

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4068 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4127

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
```

```
Db   4128 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4187

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
          |||         ||||||:::||||||   |||    :::|||        |||:::||||||
Db   4188 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4247

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
                |||   |||||||||                |||             :::   :::
Db   4248 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4307

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
          |||:::::::::|||         :::      |||   :::   ||||||   ||||||
Db   4308 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4367

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
          |||         |||::::::|||   |||   ::::::||||||||||||||     :::
Db   4368 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4427

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
             :::   |||   ||||||   |||   |||:::   :::         ||||||
Db   4428 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4487

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
          |||   ||||||||||||:::   |||      ||||||||||||:::      ||||||
Db   4488 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4547

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
          ||||||                :::                ||||||   ||||||   |||
Db   4548 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4607

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
          |||:::   ||||||:::|||   |||:::||||||:::            :::      |||
Db   4608 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4667

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
                |||||||:::||||||   :::   |||   :::   :::|||:::|||   |||
Db   4668 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4727

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
          :::         |||      :::         |||   |||   |||||||||:::   |||
Db   4728 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTA 4781

Qy   1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
          :::   |||         |||||||||||||||:::||||||:::|||:::|||   |||:::
Db   4782 GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATC 4841

Qy   1511 IleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThrHisAla 1529
          ||||||||||||||||||||||||   ||||||   |||||||||   |||
Db   4842 ATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACG 4901

Qy   1530 TyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCys 1549
             :::   |||||||||            :::   ||||||::::::      :::
Db   4902 TCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAA 4961

Qy   1550 TrpThrLys---------------MetLysAspAla-LysMetLysHisProValProGl 1564
                |||                      ||| :::|||||||||   :::   ||
Db   4962 ACGGAGAAAGAAGGAGAGGACCCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCA 5021

Qy   1564 nProArgGlu---------------------AspLeuGlnValLeuThrValLeuGluAl 1577
          ||||:::                        |||      :::|||         :::
Db   5022 GCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGAC 5081

Qy   1577 aTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
                :::|||:::|||            |||:::   |||   |||:::
Db   5082 ATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5133
```

Search completed: December 13, 2002, 06:06:25
Job time : 6800 secs

```
                     GenCore version 5.1.3
           Copyright (c) 1993 - 2002  Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on:         December 13, 2002, 00:45:36 ; Search time 1677 Seconds
                                           (without alignments)
                                           692.852 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       422006 seqs, 364007625 residues

Total number of hits satisfying chosen parameters:      844012

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09902432/runat_12122002_141453_25888/app_query.fasta_1.1735
-DB=Pending_Patents_NA_New -QFMT=fastap -SUFFIX=rnpn -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09902432_@CGN_1_1_156_@runat_12122002_141453_25888 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Pending_Patents_NA_New:*
                1:  /cgn2_6/ptodata/1/pna/PCT_NEW_COMB.seq:*
                2:  /cgn2_6/ptodata/1/pna/US06_NEW_COMB.seq:*
                3:  /cgn2_6/ptodata/1/pna/US07_NEW_COMB.seq:*
                4:  /cgn2_6/ptodata/1/pna/US08_NEW_COMB.seq:*
                5:  /cgn2_6/ptodata/1/pna/US09_NEW_COMB.seq:*
                6:  /cgn2_6/ptodata/1/pna/US10_NEW_COMB.seq:*
                7:  /cgn2_6/ptodata/1/pna/US60_NEW_COMB.seq:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8030	99.5	5236	6	US-10-152-319A-2063	Sequence 2063, Ap
2	4341.5	53.8	5931	5	US-09-724-676-12735	Sequence 12735, A
3	4341.5	53.8	5931	5	US-09-724-676A-12735	Sequence 12735, A
4	4300.5	53.3	6608	1	PCT-US02-04915-43	Sequence 43, Appl
5	4104.5	50.8	5773	5	US-09-724-676-12736	Sequence 12736, A
6	4104.5	50.8	5773	5	US-09-724-676A-12736	Sequence 12736, A
7	492	6.1	9161	6	US-10-133-937-37	Sequence 37, Appl
8	431.5	5.3	7997	6	US-10-203-138A-6112	Sequence 6112, Ap
9	427	5.3	15231	6	US-10-152-319A-1780	Sequence 1780, Ap
10	418	5.2	14335	5	US-09-724-676-25032	Sequence 25032, A
11	418	5.2	14335	5	US-09-724-676A-25032	Sequence 25032, A
12	415	5.1	8161	5	US-09-724-676-7257	Sequence 7257, Ap
13	415	5.1	8161	5	US-09-724-676A-7257	Sequence 7257, Ap
14	415	5.1	8189	5	US-09-724-676-7240	Sequence 7240, Ap
15	415	5.1	8189	5	US-09-724-676A-7240	Sequence 7240, Ap
16	415	5.1	9386	5	US-09-724-676-7256	Sequence 7256, Ap
17	415	5.1	9386	5	US-09-724-676A-7256	Sequence 7256, Ap
18	415	5.1	9414	5	US-09-724-676-7239	Sequence 7239, Ap
19	415	5.1	9414	5	US-09-724-676A-7239	Sequence 7239, Ap
20	415	5.1	9466	5	US-09-724-676-7254	Sequence 7254, Ap
21	415	5.1	9466	5	US-09-724-676A-7254	Sequence 7254, Ap
22	415	5.1	9494	5	US-09-724-676-7237	Sequence 7237, Ap
23	415	5.1	9494	5	US-09-724-676A-7237	Sequence 7237, Ap
24	415	5.1	10031	5	US-09-724-676-7255	Sequence 7255, Ap
25	415	5.1	10031	5	US-09-724-676A-7255	Sequence 7255, Ap
26	415	5.1	10059	5	US-09-724-676-7238	Sequence 7238, Ap
27	415	5.1	10059	5	US-09-724-676A-7238	Sequence 7238, Ap
28	415	5.1	10111	5	US-09-724-676-7253	Sequence 7253, Ap
29	415	5.1	10111	5	US-09-724-676A-7253	Sequence 7253, Ap
30	415	5.1	10139	5	US-09-724-676-7236	Sequence 7236, Ap
31	415	5.1	10139	5	US-09-724-676A-7236	Sequence 7236, Ap
32	415	5.1	10356	5	US-09-724-676-7252	Sequence 7252, Ap
33	415	5.1	10356	5	US-09-724-676A-7252	Sequence 7252, Ap
34	415	5.1	10384	5	US-09-724-676-7235	Sequence 7235, Ap
35	415	5.1	10384	5	US-09-724-676A-7235	Sequence 7235, Ap
36	413	5.1	5293	5	US-09-724-676-7107	Sequence 7107, Ap
37	413	5.1	5293	5	US-09-724-676A-7107	Sequence 7107, Ap
38	413	5.1	5626	5	US-09-724-676-7083	Sequence 7083, Ap
39	413	5.1	5626	5	US-09-724-676A-7083	Sequence 7083, Ap
40	412	5.1	5329	5	US-09-724-676-7104	Sequence 7104, Ap
41	412	5.1	5329	5	US-09-724-676A-7104	Sequence 7104, Ap
42	412	5.1	5662	5	US-09-724-676-7054	Sequence 7054, Ap
43	412	5.1	5662	5	US-09-724-676A-7054	Sequence 7054, Ap
44	409	5.1	5329	5	US-09-724-676-7110	Sequence 7110, Ap
45	409	5.1	5329	5	US-09-724-676A-7110	Sequence 7110, Ap

ALIGNMENTS

```
RESULT 1
US-10-152-319A-2063
; Sequence 2063, Application US/10152319A
; GENERAL INFORMATION:
;  APPLICANT: Mendrick, Donna
;  APPLICANT:  Porter, Mark
;  APPLICANT:  Johnson, Kory
;  APPLICANT:  Higgs, Brandon
;  APPLICANT:  Castle, Arthur
;  APPLICANT:  Elashoff, Michael
;  TITLE OF INVENTION: Molecular Toxicology Modeling
;  FILE REFERENCE: 44921-5089-US
;  CURRENT APPLICATION NUMBER: US/10/152,319A
;  CURRENT FILING DATE:  2002-05-22
;  PRIOR APPLICATION NUMBER: US 60/292,335
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/297,523
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,925
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,810
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,807
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/303,808
;  PRIOR FILING DATE: 2001-07-10
;  PRIOR APPLICATION NUMBER: US 60/315,047
;  PRIOR FILING DATE: 2001-08-28
;  PRIOR APPLICATION NUMBER: US 60/324,928
;  PRIOR FILING DATE: 2001-09-27
;  PRIOR APPLICATION NUMBER: US 60/330,867
;  PRIOR FILING DATE: 2001-11-01
;  PRIOR APPLICATION NUMBER: US 60/330,462
;  PRIOR FILING DATE: 2001-10-22
;  Remaining Prior Application data removed - See File Wrapper or PALM.
```

```
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2063
;   LENGTH: 5236
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. NM_057103
US-10-152-319A-2063

Alignment Scores:
Pred. No.:              0          Length:        5236
Score:                  8030.00    Matches:       1589
Percent Similarity:     99.75%     Conservative:  3
Best Local Similarity:  99.56%     Mismatches:    4
Query Match:            99.47%     Indels:        0
DB:                     6          Gaps:          0

US-09-902-432-4 (1-1596) x US-10-152-319A-2063 (1-5236)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92

Qy     21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy     41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy     61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy     81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy    101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy    341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy    361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy    381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy    401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy    421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy    441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy    461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy    481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532

Qy    501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592

Qy    521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652

Qy    541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy    561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy    581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy    601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy    621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy    641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy    661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy    681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132

Qy    701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy    741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy    761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2433 AGGGCAGATGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
          ||||||||||||:::|||||||||||||||||||||||||||||:::|||||||||||||
Db   2493 GAGGACGACCCTGATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGG 2552

Qy    841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
          |||||||||||||||||||||||      |||||||||||||||||||||||||||||||
Db   2553 GAGAAGATGGAAGCCCAGGGGAATGCGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612

Qy    861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy    881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy    901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy    921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy    941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy    961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy    981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy   1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy   1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy   1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212

Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812

Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy   1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy   1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy   1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112
```

```
Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
Db   4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
Db   4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
Db   4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy   1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
Db   4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy   1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
Db   4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy   1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
Db   4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy   1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
Db   4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy   1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
Db   4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy   1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
Db   4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCAG 4652

Qy   1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
Db   4653 GCTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACAC 4712

Qy   1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
Db   4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772

Qy   1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
Db   4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820
```

RESULT 2
US-09-724-676-12735
; Sequence 12735, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12735
;   LENGTH: 5931
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-12735

Alignment Scores:			
Pred. No.:	3.65e-172	Length:	5931
Score:	4341.50	Matches:	970
Percent Similarity:	69.13%	Conservative:	190
Best Local Similarity:	57.81%	Mismatches:	408
Query Match:	53.78%	Indels:	110
DB:	5	Gaps:	33

US-09-902-432-4 (1-1596) x US-09-724-676-12735 (1-5931)

```
Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln---------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACAGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGACACCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 671

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    672 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 731

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    732 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 791

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    792 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 842

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    843 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 896

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    897 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 956

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    957 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1013

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1014 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1073

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1074 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1133

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1134 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1193

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu-------------- 337
Db   1194 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1253
```

```
Qy    338 -----------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1254 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1313

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1314 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1373

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1374 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1433

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1434 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1484

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1485 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1529

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1530 TTGGTTGAAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1589

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1590 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1649

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1650 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1709

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1710 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1769

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1770 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1826

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1827 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1886

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1887 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1946

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1947 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2006

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2007 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2066

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2067 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2126

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2127 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2186

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2187 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2246

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2247 GGATCATCCAAGAAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2306

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2307 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2366

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2367 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2426

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2427 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2486

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2487 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2546

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2547 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2606

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2607 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2666

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2667 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2726

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2727 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2786

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2787 CAAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2846

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2847 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2906

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2907 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2966

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2967 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3026

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3027 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3086

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3087 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3146

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3147 GGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3206

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3207 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3266

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3267 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3326

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3327 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3386

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
```

```
             |||  |||||||||                    :::  |||    ||||||
Db    3387 CTTCAGCCTGTGCAGAGAGCA-----------------GAGGCAGAAAGACCAGAAGAG 3428

Qy    1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
              :::|||          :::||||||  ||||||:::  |||          |||
Db    3429 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3488

Qy    1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
           ||||||     |||        ||||||          ||||||   ||||||||||||
Db    3489 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3548

Qy    1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
              |||      |||:::||||||     ::::::            |||   ||||||
Db    3549 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3608

Qy    1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
           |||                  :::|||  |||::::::   ||||||||||||:::
Db    3609 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3668

Qy    1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
           |||:::|||   ||||||   |||||||||||||||:::||||||||||||:::||||||
Db    3669 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3728

Qy    1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
              |||             ||||||:::||||||   ::::::      :::
Db    3729 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3788

Qy    1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
              |||:::        ||||||         ||||||   ||||||      ||||||||||||
Db    3789 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3848

Qy    1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
              |||          |||::::::         |||||||||      |||:::
Db    3849 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3905

Qy    1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
                 |||   ||||||  |||   :::|||::::::         |||    :::|||
Db    3906 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3965

Qy    1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
           :::|||||||||         |||      |||   |||         ||||||   |||
Db    3966 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4019

Qy    1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
                          |||      |||      :::   :::|||          :::
Db    4020 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4079

Qy    1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
           :::|||      :::|||            |||            |||   |||::::::
Db    4080 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4139

Qy    1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
              ::::::   |||:::||||||                     |||      |||   |||
Db    4140 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4199

Qy    1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
           |||         ||||||:::||||||   |||   :::|||      |||:::||||||
Db    4200 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4259

Qy    1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
                  |||   |||||||||               |||            :::   :::
Db    4260 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4319

Qy    1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
           |||:::::::::|||             :::   |||   :::   ||||||   ||||||
Db    4320 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4379

Qy    1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
           |||               |||::::::|||   |||      ::::::||||||||||||   :::
Db    4380 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4439

Qy    1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
              :::   |||   ||||||   |||   |||:::   :::         ||||||
Db    4440 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4499

Qy    1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
           |||   |||||||||||||:::   |||         ||||||||||||:::      ||||||
Db    4500 ATTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4559

Qy    1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
           ||||||                 :::                  ||||||   ||||||   |||
Db    4560 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4619

Qy    1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
           |||:::   ||||||:::|||   |||:::||||||:::               :::      |||
Db    4620 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4679

Qy    1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
                    ||||||:::||||||   :::   |||   :::   :::|||:::|||   |||
Db    4680 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4739

Qy    1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
           :::         |||      :::            |||   |||   |||||||||:::   |||
Db    4740 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTA 4793

Qy    1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
           :::   |||      |||||||||||||||:::||||||:::|||:::|||   |||:::
Db    4794 GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATC 4853

Qy    1511 IleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThrHisAla 1529
           |||||||||||||||||||||      ||||||   |||||||||      |||
Db    4854 ATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACG 4913

Qy    1530 TyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCys 1549
              :::   |||||||||         :::   ||||||::::::         :::
Db    4914 TCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAA 4973

Qy    1550 TrpThrLys--------------MetLysAspAla-LysMetLysHisProValProGl 1564
              |||                       |||  :::|||||||||      :::   ||
Db    4974 ACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCA 5033

Qy    1564 nProArgGlu--------------------AspLeuGlnValLeuThrValLeuGluAl 1577
           ||||:::                        |||      :::|||         :::
Db    5034 GCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGAC 5093

Qy    1577 aTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
                 :::|||:::|||               |||:::   |||   |||:::
Db    5094 ATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5145
```

RESULT 3
US-09-724-676A-12735
; Sequence 12735, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12735
; LENGTH: 5931
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-12735

Alignment Scores:

Pred. No.:	3.65e-172	Length:	5931
Score:	4341.50	Matches:	970
Percent Similarity:	69.13%	Conservative:	190
Best Local Similarity:	57.81%	Mismatches:	408
Query Match:	53.78%	Indels:	110
DB:	5	Gaps:	33

US-09-902-432-4 (1-1596) x US-09-724-676A-12735 (1-5931)

```
Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln---------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACAGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGACACCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 671

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    672 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 731

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    732 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 791

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    792 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 842

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    843 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 896

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    897 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 956

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    957 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1013

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1014 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1073

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1074 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1133

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1134 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1193

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1194 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1253

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1254 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1313

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1314 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1373

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1374 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1433

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1434 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1484

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1485 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1529

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1530 TTGGTTGAAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1589

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1590 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1649

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1650 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1709

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1710 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1769

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1770 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1826

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1827 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1886

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1887 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1946

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1947 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2006

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2007 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2066

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2067 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2126

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2127 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2186

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2187 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2246

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2247 GGATCATCCAAGAAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2306

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
```

```
Db   2307 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2366

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2367 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2426

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2427 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2486

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2487 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAGCGAAGACTCCATAGCT 2546

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2547 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2606

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2607 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2666

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2667 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2726

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2727 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2786

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2787 CAAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2846

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2847 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2906

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2907 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2966

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2967 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3026

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3027 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3086

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3087 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3146

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3147 GGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3206

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3207 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3266

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3267 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3326

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3327 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3386

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3387 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3428

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3429 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3488

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3489 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3548

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3549 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3608

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3609 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3668

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3669 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3728

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3729 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3788

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3789 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3848

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3849 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3905

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3906 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3965

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3966 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4019

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4020 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4079

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4080 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4139

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4140 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4199

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4200 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4259

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4260 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4319

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4320 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4379

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4380 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4439

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4440 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4499
```

```
Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4500 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4559

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4560 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4619

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4620 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4679

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4680 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4739

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4740 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTA 4793

Qy   1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
Db   4794 GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATC 4853

Qy   1511 IleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThrHisAla 1529
Db   4854 ATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACG 4913

Qy   1530 TyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCys 1549
Db   4914 TCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAA 4973

Qy   1550 TrpThrLys--------------MetLysAspAla-LysMetLysHisProValProGl 1564
Db   4974 ACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCA 5033

Qy   1564 nProArgGlu--------------------AspLeuGlnValLeuThrValLeuGluAl 1577
Db   5034 GCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGAC 5093

Qy   1577 aTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   5094 ATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5145
```

RESULT 4

PCT-US02-04915-43

```
; Sequence 43, Application PC/TUS0204915
; GENERAL INFORMATION:
;  APPLICANT: Murray, Richard
;  APPLICANT:  Glynne, Richard
;  APPLICANT:  Watson, Susan R.
;  APPLICANT:  Aziz, Natasha
;  APPLICANT:  Eos Biotechnology, Inc.
;  TITLE OF INVENTION: Methods of Diagnosis of Angiogenesis, Compositions and
;  TITLE OF INVENTION:  Methods of Screening for Angiogenesis Modulators
;  FILE REFERENCE: 018501-006200PC
;  CURRENT APPLICATION NUMBER: PCT/US02/04915
;  CURRENT FILING DATE:  2002-02-14
;  PRIOR APPLICATION NUMBER: US 09/784,356
;  PRIOR FILING DATE: 2001-02-14
;  PRIOR APPLICATION NUMBER: US 09/791,390
;  PRIOR FILING DATE: 2001-02-22
;  PRIOR APPLICATION NUMBER: US 60/285,475
;  PRIOR FILING DATE: 2002-04-19
;  PRIOR APPLICATION NUMBER: US 60/310,025
;  PRIOR FILING DATE: 2001-08-03
;  PRIOR APPLICATION NUMBER: US 60/350,666
;  PRIOR FILING DATE: 2001-11-13
;  PRIOR APPLICATION NUMBER: US 60/334,244
;  PRIOR FILING DATE: 2001-11-29
;  NUMBER OF SEQ ID NOS: 230
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 43
;   LENGTH: 6608
;   TYPE: DNA
;   ORGANISM: Homo sapiens
PCT-US02-04915-43
```

```
Alignment Scores:
Pred. No.:               1.99e-170     Length:         6608
Score:                   4300.50       Matches:        967
Percent Similarity:      68.81%        Conservative:   189
Best Local Similarity:   57.56%        Mismatches:     408
Query Match:             53.27%        Indels:         116
DB:                      1             Gaps:           36
```

US-09-902-432-4 (1-1596) x PCT-US02-04915-43 (1-6608)

```
Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln---------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
```

```
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203
```

```
Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3845

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3846 AGGCCTCCAGCA---CCTTCCAGTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3902

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3903 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3962

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3963 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4016

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4017 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4076

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4077 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4136

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4137 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4196

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4197 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4256

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4257 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4316

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4317 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4376

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4377 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4436

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4437 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4496

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4497 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4556

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4557 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4616

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4617 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4676

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4677 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4736

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4737 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4787

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
Db   4788 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4841

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
Db   4842 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4901

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
Db   4902 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4961

Qy   1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
Db   4962 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5021

Qy   1562 lProGlnProArgGlu---------------------AspLeuGlnValLeuThrValLe 1575
Db   5022 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5081

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   5082 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5139
```

RESULT 5
US-09-724-676-12736
; Sequence 12736, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12736
;   LENGTH: 5773
;   TYPE: DNA

```
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (91)..(91)
;   OTHER INFORMATION: n is a,c,g, or t
US-09-724-676-12736

Alignment Scores:
Pred. No.:                2.33e-162      Length:          5773
Score:                    4104.50        Matches:         912
Percent Similarity:       69.40%         Conservative:    179
Best Local Similarity:    58.02%         Mismatches:      389
Query Match:              50.84%         Indels:          92
DB:                       5              Gaps:            29

US-09-902-432-4 (1-1596) x US-09-724-676-12736 (1-5773)

Qy       89 ValGlyGlnArgGluSerGluAspValArgGluLysAspArgValGluGluMetAlaAla 108
            ||||||||||:::||||||||||||    |||:::|||       :::||||||||
Db      352 GTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACT 411

Qy      109 AsnSerThrAlaValGluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIle 128
               |||     |||   ||||||||   |||||||||||||||||     ||||||||
Db      412 AAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATC 471

Qy      129 GluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnProAlaGluSerGln 148
            |||||||||||:::||||||:::|||:::||||||:::     ||||||   ||||||||
Db      472 GAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAG 531

Qy      149 AlaAsnAspValGlyPheLysLysValPheLysPheValGlyPheLysPheThrValLys 168
            ||||||||:::|||||||||||||||||||||||||||||||||||||||||||||||||
Db      532 GCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAA 591

Qy      169 LysAspLysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGly 188
            |||||||||    ||||||   ||||||||||||||||||||||||||||||||||||||
Db      592 AAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGG 651

Qy      189 GluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAla 208
            |||||||||         ||||||||||||||||||:::||||||:::
Db      652 GAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGACCCCAGCCTT------GGG 696

Qy      209 ValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGly 228
               ||||||:::||||||||||||||||||    ||||||||||||||||||   |||
Db      697 GCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAG 756

Qy      229 ThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAla 248
            ||||||||:::||||||||||         ||||||         |||||||||   ||||||
Db      757 ACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCA 816

Qy      249 AlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSer 268
                  ||||||   |||:::||||||||||||||||||||||||||||:::||||||
Db      817 GTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCT 873

Qy      269 ProGluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePhe 288
             |||||||||:::|||||||||    |||||||||    |||:::||||||||||||||
Db      874 GCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTC 933

Qy      289 ThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAsp 308
            |||   ||||||||||||||||||||||||||||:::|||   ||||||||||:::
Db      934 ACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAA 993

Qy      309 LeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLys 328
            :::|||    :::||||||:::||||||||||||   |||||||||||   ||||||
Db      994 GTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGAC 1053

Qy      329 GluLysThrGluProAlaSerGluGlu---------------------------GlnGlu 339
               |||   |||   ||||||||||:::                          ||||||
Db     1054 GGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAG 1113

Qy      340 ProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuPro 359
            |||||||||         :::   |||||||||||:::|||||||||||||||||||||
Db     1114 CCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCC 1173

Qy      360 LeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAla 379
               |||:::||||||      :::   |||||||||||||   ||||||||||||
Db     1174 TCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCG 1233

Qy      380 ThrGluValPheAspGluLysMetGluAlaHisGln---GluValValAlaGluValHis 398
            ||||||||||||||||||||:::|||   ||||||   ||||||||||||||||||
Db     1234 ACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCAC 1293

Qy      399 ValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGly 418
            ||||||||||||||||:::   |||||||||
Db     1294 GTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG-----------------------AAA 1329

Qy      419 ValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGlu 438
                  ||||||   |||   |||:::|||   |||:::|||   |||
Db     1330 ACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCA 1389

Qy      439 ValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSer 458
               |||||||||||||||||||||:::|||||||:::|||   :::|||   ||||||||
Db     1390 GAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCC 1449

Qy      459 GlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHis 478
            |||   |||   ||||||      |||||||||||||:::||||||   |||   |||
Db     1450 GGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCC 1509

Qy      479 ProGluGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGln 498
            ||||||||:::||||||||||||||||||||||||||||||||||||:::|||||||||
Db     1510 CCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAG 1569

Qy      499 GlySerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLys 518
            |||||||||||||||||||||:::|||:::||||||||||||||||||||||||||||||
Db     1570 GGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAA 1629

Qy      519 GlnLysGlyLysArgGlyGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHis 538
            ||||||||||||      ||||||||||||||||   |||||||:::         :::
Db     1630 CAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCA 1686

Qy      539 ThrGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGlu 558
               :::||||||:::|||   :::||||||||||||||||||||||||||||||||||
Db     1687 GCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAG 1746

Qy      559 GluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGlu 578
            ||||||||||||   ||||||||||||||||   |||      ||||||||||||
Db     1747 GAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAA 1806

Qy      579 AlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAla 598
            ||||||||||||   ||||||||||||||||||||||||||||:::||||||||||||
Db     1807 GCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCA 1866

Qy      599 SerPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLys 618
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1867 TCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAA 1926

Qy      619 GluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSer 638
            |||:::||||||:::||||||||||||||||||||||||||||:::||||||   |||
Db     1927 GAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCT 1986

Qy      639 GluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArg 658
            ||||||||:::|||:::|||         ||||||   ||||||||||||||||||
Db     1987 GAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGC 2046

Qy      659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
            :::|||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2047 AAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGA 2106

Qy      679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
            ||||||:::   ||||||||||||:::||||||||||||:::      :::|||||||||
Db     2107 GCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCAC 2166

Qy      699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
               :::|||:::||||||   ||||||||||||   |||||||||    :::   |||
Db     2167 CAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGT 2226

Qy      719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
```

```
Db   2227 TCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACC 2286

Qy    739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLys 758
Db   2287 GAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAA 2346

Qy    759 SerLysSerLysLeuGluGluLysAlaGluAsp------------SerSerValGluGln 774
Db   2347 TCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACAT 2406

Qy    775 LeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIle 794
Db   2407 TCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATT 2466

Qy    795 ProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSer 814
Db   2467 CCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCA 2526

Qy    815 GlyProValGluIleAsnGluAspAspProAsnValProAlaValValProLeuSerGlu 834
Db   2527 GGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAG 2586

Qy    835 TyrAsnAlaValGluArgGluLysMetGlu---------AlaGlnGlyAsnThrGluLeu 851
Db   2587 TATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAG 2646

Qy    852 ProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThr 871
Db   2647 CCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATG 2706

Qy    872 ValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSerPro 891
Db   2707 ATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCT 2766

Qy    892 SerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetPro 911
Db   2767 TCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTG 2826

Qy    912 ProValGluGluValThrGluLysAspIleIleAla---GluGluThrProValLeuThr 930
Db   2827 TTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACT 2886

Qy    931 GlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAspPhe 950
Db   2887 GAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTG 2946

Qy    951 ThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThr 970
Db   2947 ACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACC 3006

Qy    971 GluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThrAsp 990
Db   3007 GAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGAC 3066

Qy    991 SerProAspThrThrGluGluAlaThrProValGlnGluValGluSerGlyValLeuAsp 1010
Db   3067 TCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGAC 3126

Qy   1011 ThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLys 1030
Db   3127 ATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAA 3186

Qy   1031 GluGluSerGlnValProAlaThr--------------------GlnThrValGlnArg 1043
Db   3187 GAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGA 3246

Qy   1044 ThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAla 1063
Db   3247 GCA-----------------GAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGT 3288

Qy   1064 SerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHis 1083
Db   3289 CTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCT 3348

Qy   1084 LeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGlu---ValProGlu 1102
Db   3349 TTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAA 3408

Qy   1103 ValThrAlaAspValAspHis---------ValAlaThrCysGlnVal----------- 1115
Db   3409 GTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCT 3468

Qy   1116 ---IleLysLeuGlnGlnLeu---MetGluGlnAlaValAlaProGluSerSerGluThr 1133
Db   3469 GGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACC 3528

Qy   1134 LeuThrAspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGly 1153
Db   3529 CCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACA 3588

Qy   1154 ThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGln 1173
Db   3589 ACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAG 3648

Qy   1174 SerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuPro 1193
Db   3649 TCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCA---CCT 3705

Qy   1194 AsnAsnValProAlaGlnGluGluHisGlyGluGluProGly---ArgAspValLeuGlu 1212
Db   3706 TCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAG 3765

Qy   1213 ProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluValGly 1232
Db   3766 CATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACT 3825

Qy   1233 GlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys---------------Glu 1247
Db   3826 CAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAAGACGTACCATTTTTCGAA 3879

Qy   1248 GluGlnGluValPheValHisSerGly------ProAsnSerGlnLysAlaAlaAspVal 1265
Db   3880 GGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGGGAAAAGGTCACTGAAGTT 3939

Qy   1266 ThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluVal 1285
Db   3940 GCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAGGATGATGCTCTTGAACTG 3999

Qy   1286 GlnSer---------------LeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
Db   4000 CAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAA 4059

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGly---GluGlnGluThr 1319
Db   4060 AGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAGAAGCTTGAGCACGAAACA 4119

Qy   1320 AlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSer 1339
Db   4120 GCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAGACAGTGAATGTGCCCATC 4179

Qy   1340 SerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer---ProSerLeuProAspGln 1358
Db   4180 ATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAA 4239

Qy   1359 AspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThr 1378
Db   4240 GAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAGGCATCATTCACTCTAACA 4299

Qy   1379 AlaGluAlaValGluLysValIle------GluThrValValIleSerGluThrGlyGlu 1396
Db   4300 GCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAACATTTTAGAAACAGGTGAA 4359

Qy   1397 SerProGluCysValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGly 1416
Db   4360 ACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAATCCTCTGAAAAAAATGAA 4419
```

```
Qy   1417 HisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGlu 1436
Db   4420 GACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGGCCCGACTGTCAGGCAAAA 4479

Qy   1437 SerIleProIleIleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGly 1456
Db   4480 TCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGA 4539

Qy   1457 GluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyPro 1476
Db   4540 GAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGATGAGCAGGTTGCTTGCCAG 4599

Qy   1477 AspAlaAspGlyLysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIle 1496
Db   4600 GAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTAGAGCCTGAAAATGGGATT 4653

Qy   1497 LeuGluLeuGluSerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAsp 1516
Db   4654 TTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATCATCCAGACAGCCGTTGAC 4713

Qy   1517 GlnPheAlaArgThr---GluThrAlaProGluThrHisAlaTyrAspSerGlnThrGln 1535
Db   4714 CAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACGTCTGAGTTACAGACACAA 4773

Qy   1536 ValProAlaCysArgLeuAspSerArgGluProAsnArgCysTrpThrLys--------- 1552
Db   4774 GCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAAACGGAGAAAGAAGGAGAG 4833

Qy   1553 ------MetLysAspAla-LysMetLysHisProValProGlnProArgGlu-------- 1567
Db   4834 GAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCAGCCAAAGAGGAGTCAGAG 4893

Qy   1568 -------------AspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLy 1583
Db   4894 TCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGACATGAGTGAAGCCTCAGAA 4953

Qy   1583 sCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   4954 AAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 4987
```

RESULT 6
US-09-724-676A-12736
; Sequence 12736, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12736
; LENGTH: 5773
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (91)..(91)
; OTHER INFORMATION: n is a,c,g, or t
US-09-724-676A-12736

Alignment Scores:

Pred. No.:	2.33e-162	Length:	5773
Score:	4104.50	Matches:	912
Percent Similarity:	69.40%	Conservative:	179
Best Local Similarity:	58.02%	Mismatches:	389
Query Match:	50.84%	Indels:	92
DB:	5	Gaps:	29

US-09-902-432-4 (1-1596) x US-09-724-676A-12736 (1-5773)

```
Qy     89 ValGlyGlnArgGluSerGluAspValArgGluLysAspArgValGluGluMetAlaAla 108
Db    352 GTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACT 411

Qy    109 AsnSerThrAlaValGluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIle 128
Db    412 AAGTCAGCGGTTGTTCACGACATCACAGATGATGGCAGGAGGAGACACCCGAAATAATC 471

Qy    129 GluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnProAlaGluSerGln 148
Db    472 GAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAG 531

Qy    149 AlaAsnAspValGlyPheLysLysValPheLysPheValGlyPheLysPheThrValLys 168
Db    532 GCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAA 591

Qy    169 LysAspLysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGly 188
Db    592 AAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGG 651

Qy    189 GluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAla 208
Db    652 GAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGACCCCAGCCTT------GGG 696

Qy    209 ValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGly 228
Db    697 GCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAG 756

Qy    229 ThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAla 248
Db    757 ACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCA 816

Qy    249 AlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSer 268
Db    817 GTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCT 873

Qy    269 ProGluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePhe 288
Db    874 GCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTC 933

Qy    289 ThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAsp 308
Db    934 ACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAA 993

Qy    309 LeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLys 328
Db    994 GTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGAC 1053

Qy    329 GluLysThrGluProAlaSerGluGlu--------------------------GlnGlu 339
Db   1054 GGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAG 1113

Qy    340 ProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuPro 359
Db   1114 CCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCC 1173

Qy    360 LeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAla 379
Db   1174 TCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCG 1233

Qy    380 ThrGluValPheAspGluLysMetGluAlaHisGln---GluValValAlaGluValHis 398
Db   1234 ACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCAC 1293

Qy    399 ValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGly 418
Db   1294 GTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG-----------------------AAA 1329

Qy    419 ValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGlu 438
Db   1330 ACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCA 1389

Qy    439 ValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSer 458
```

```
Db   1390 GAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCC 1449

Qy    459 GlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHis 478
Db   1450 GGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCC 1509

Qy    479 ProGluGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGln 498
Db   1510 CCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAG 1569

Qy    499 GlySerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLys 518
Db   1570 GGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAA 1629

Qy    519 GlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHis 538
Db   1630 CAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCA 1686

Qy    539 ThrGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGlu 558
Db   1687 GCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAG 1746

Qy    559 GluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGlu 578
Db   1747 GAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAA 1806

Qy    579 AlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAla 598
Db   1807 GCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCA 1866

Qy    599 SerPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLys 618
Db   1867 TCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAA 1926

Qy    619 GluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSer 638
Db   1927 GAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCT 1986

Qy    639 GluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArg 658
Db   1987 GAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGC 2046

Qy    659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
Db   2047 AAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGA 2106

Qy    679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
Db   2107 GCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCAC 2166

Qy    699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
Db   2167 CAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGT 2226

Qy    719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
Db   2227 TCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACC 2286

Qy    739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLys 758
Db   2287 GAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAA 2346

Qy    759 SerLysSerLysLeuGluGluLysAlaGluAsp------------SerSerValGluGln 774
Db   2347 TCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACAT 2406

Qy    775 LeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIle 794
Db   2407 TCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATT 2466

Qy    795 ProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSer 814
Db   2467 CCTGGACGAAGGAAGAAAAGGCCAGATGGAAACAAGAACAAGCCCCTGTTGAAGACGCA 2526

Qy    815 GlyProValGluIleAsnGluAspAspProAsnValProAlaValValProLeuSerGlu 834
Db   2527 GGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAG 2586

Qy    835 TyrAsnAlaValGluArgGluLysMetGlu---------AlaGlnGlyAsnThrGluLeu 851
Db   2587 TATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAG 2646

Qy    852 ProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThr 871
Db   2647 CCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATG 2706

Qy    872 ValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSerPro 891
Db   2707 ATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCT 2766

Qy    892 SerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetPro 911
Db   2767 TCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTG 2826

Qy    912 ProValGluGluValThrGluLysAspIleIleAla---GluGluThrProValLeuThr 930
Db   2827 TTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACT 2886

Qy    931 GlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAspPhe 950
Db   2887 GAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTG 2946

Qy    951 ThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThr 970
Db   2947 ACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACC 3006

Qy    971 GluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThrAsp 990
Db   3007 GAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGAC 3066

Qy    991 SerProAspThrThrGluGluAlaThrProValGlnGluValGluSerGlyValLeuAsp 1010
Db   3067 TCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGAC 3126

Qy   1011 ThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLys 1030
Db   3127 ATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAA 3186

Qy   1031 GluGluSerGlnValProAlaThr--------------------GlnThrValGlnArg 1043
Db   3187 GAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGA 3246

Qy   1044 ThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAla 1063
Db   3247 GCA-----------------GAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGT 3288

Qy   1064 SerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHis 1083
Db   3289 CTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCT 3348

Qy   1084 LeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGlu---ValProGlu 1102
Db   3349 TTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAA 3408

Qy   1103 ValThrAlaAspValAspHis---------ValAlaThrCysGlnVal------------ 1115
Db   3409 GTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCT 3468

Qy   1116 ---IleLysLeuGlnGlnLeu---MetGluGlnAlaValAlaProGluSerSerGluThr 1133
Db   3469 GGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACC 3528

Qy   1134 LeuThrAspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGly 1153
Db   3529 CCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACA 3588
```

```
Qy   1154 ThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGln 1173
Db   3589 ACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAG 3648

Qy   1174 SerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuPro 1193
Db   3649 TCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCA---CCT 3705

Qy   1194 AsnAsnValProAlaGlnGluGluHisGlyGluGluProGly---ArgAspValLeuGlu 1212
Db   3706 TCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAG 3765

Qy   1213 ProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluValGly 1232
Db   3766 CATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACT 3825

Qy   1233 GlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys---------------Glu 1247
Db   3826 CAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAAGACGTACCATTTTTCGAA 3879

Qy   1248 GluGlnGluValPheValHisSerGly------ProAsnSerGlnLysAlaAlaAspVal 1265
Db   3880 GGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGGGAAAAGGTCACTGAAGTT 3939

Qy   1266 ThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluVal 1285
Db   3940 GCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAGGATGATGCTCTTGAACTG 3999

Qy   1286 GlnSer---------------LeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
Db   4000 CAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAA 4059

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGly---GluGlnGluThr 1319
Db   4060 AGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAGAAGCTTGAGCACGAAACA 4119

Qy   1320 AlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSer 1339
Db   4120 GCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAGACAGTGAATGTGCCCATC 4179

Qy   1340 SerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer---ProSerLeuProAspGln 1358
Db   4180 ATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAA 4239

Qy   1359 AspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThr 1378
Db   4240 GAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAGGCATCATTCACTCTAACA 4299

Qy   1379 AlaGluAlaValGluLysValIle------GluThrValValIleSerGluThrGlyGlu 1396
Db   4300 GCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAACATTTTAGAAACAGGTGAA 4359

Qy   1397 SerProGluCysValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGly 1416
Db   4360 ACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAATCCTCTGAAAAAAATGAA 4419

Qy   1417 HisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGlu 1436
Db   4420 GACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGGCCCGACTGTCAGGCAAAA 4479

Qy   1437 SerIleProIleIleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGly 1456
Db   4480 TCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGA 4539

Qy   1457 GluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyPro 1476
Db   4540 GAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGATGAGCAGGTTGCTTGCCAG 4599

Qy   1477 AspAlaAspGlyLysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIle 1496
Db   4600 GAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTAGAGCCTGAAAATGGGATT 4653

Qy   1497 LeuGluLeuGluSerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAsp 1516
Db   4654 TTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATCATCCAGACAGCCGTTGAC 4713

Qy   1517 GlnPheAlaArgThr---GluThrAlaProGluThrHisAlaTyrAspSerGlnThrGln 1535
Db   4714 CAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACGTCTGAGTTACAGACACAA 4773

Qy   1536 ValProAlaCysArgLeuAspSerArgGluProAsnArgCysTrpThrLys--------- 1552
Db   4774 GCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAAACGGAGAAAGAAGGAGAG 4833

Qy   1553 ------MetLysAspAla-LysMetLysHisProValProGlnProArgGlu-------- 1567
Db   4834 GAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCAGCCAAAGAGGAGTCAGAG 4893

Qy   1568 -------------AspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLy 1583
Db   4894 TCAACCGCAGTGGGACAAGCACATTCTGATATTTCAAAGACATGAGTGAAGCCTCAGAA 4953

Qy   1583 sCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   4954 AAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 4987

RESULT 7
US-10-133-937-37
; Sequence 37, Application US/10133937
; GENERAL INFORMATION:
;  APPLICANT: Khan, Javed
;  APPLICANT:  Ringner, Markus
;  APPLICANT:  Peterson, Carsten
;  APPLICANT:  Meltzer, Paul
;  TITLE OF INVENTION: METHODS FOR ANALYZING HIGH DIMENSIONAL DATA FOR CLASSIFYING,
;  TITLE OF INVENTION:  DIAGNOSING, PROGNOSTICATING, AND/OR PREDICTING DISEASES AND
;  TITLE OF INVENTION:  OTHER BIOLOGICAL STATES
;  FILE REFERENCE: 11613.56US01
;  CURRENT APPLICATION NUMBER: US/10/133,937
;  CURRENT FILING DATE:  2002-11-04
;  NUMBER OF SEQ ID NOS: 99
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 37
;   LENGTH: 9161
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-133-937-37

Alignment Scores:
Pred. No.:               1.41e-12      Length:          9161
Score:                   492.00        Matches:         386
Percent Similarity:      33.47%        Conservative:    269
Best Local Similarity:   19.72%        Mismatches:      714
Query Match:             6.09%         Indels:          589
DB:                      6             Gaps:            75

US-09-902-432-4 (1-1596) x US-10-133-937-37 (1-9161)

Qy     45 AlaAspProAlaThrLysLeu---------------ProGlnLysAsnGlyGlnLeuSer 59
Db   1399 GCAAACCCTGCGGAGAAAATCATCCGAGTCCTGTTTCCTGGGAACAGCACCCAGTACAAC 1458

Qy     60 SerValAsnGlyValAlaGluGlnGlyAspValHis------------------ValGln 73
Db   1459 ATCCTGGAAGGGTTGGAAAAGCTCAAACATCTAGACTTTCTGAAGCAGCCACTGGCCACC 1518

Qy     74 GluGluAsnGlnGluGlyGlnGluGluGluValValAspGluAspValGlyGlnArgGlu 93
Db   1519 CAAAAGGATCTCACTGGCCAGGTGCCCACTCCTGTGGTGAAACAAACAAAACTGAAACAG 1578

Qy     94 SerGluAspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaVal 113
Db   1579 AGGGCTGATAGCCGAGAAAGT---------CTGAAGCCAGCCGCAAAACCACTTCCTAGC 1629

Qy    114 GluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIleGluGlnIleProAla 133
```

```
Db   1630 AAATCCGTGCGCAAGGAGTCAAAAGAAGAAACCCCTGAGGTC---------------ACA 1674

Qy    134 SerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGly 153
Db   1675 AAAGTGAATCACGTGGAA---------AAGCCACCCAAAGTTGAAAGCAAAGAA------ 1719

Qy    154 PheLysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGlu 173
Db   1720 ------------------------------AAGGTAATGGTGAAAAAAGACAAGCCAGTA 1749

Qy    174 LysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAla 193
Db   1750 AAAACAGAGACCAAACCTTCAGTGACTGAAAAGGAG------------------------ 1785

Qy    194 SerValGlyAlaGlyAspHisGlnGluProSer---ValGluThrAlaValGlyGluSer 212
Db   1786 ---------GTTCCCAGCAAAGAAGAGCCATCTCCAGTGAAGCCGAGGTGGCTGAG--- 1833

Qy    213 AlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGln 232
Db   1834 ---------------------AAGCAAGCCACAGATGTCAAACCCAAAGCTGCCAAGGAG 1872

Qy    233 GluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGlu 252
Db   1873 AAGACGGTGAAAAAGGAAACAAAGGTAAAGCCTGAAGACAAGAAAGAGGAGAAAGAAAAG 1932

Qy    253 AlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerPro 272
Db   1933 CCAAAGAAAGAAGTGGCTAAAAAGGAGGACAAAACACCTATCAAGAAGGAGGAAAAACCA 1992

Qy    273 SerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrp 292
Db   1993 AAA---------AAGGAAGAGGTGAAAAAAGAAGTCAAAAAAGAGATC------------ 2031

Qy    293 AlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGluThrAla 312
Db   2032 ---------AAGAAAGAAGAGAAAAAAGAACCCAAGAAAGAGGTTAAGAAAGAAACACCG 2082

Qy    313 GluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGluLysThrGlu 332
Db   2083 CCAAAGGAAGTCAAGAAGGAAGTTAAGAAG---------------GAAGAGAAGAAGGAA 2127

Qy    333 ProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAsp 352
Db   2128 GTGAAAAAGGAAGAAAAGGAACCCAAAAAAGAAATT------AAGAAGCTCCCTAAAGAC 2181

Qy    353 TyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGlu 372
Db   2182 GCAAAGAAATCATCTACTCCTCTG---------------------------TCTGAAGCA 2214

Qy    373 GluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGlu 392
Db   2215 AAAAAACCAGCTGCTTTAAAACCAAAAGTACCCAAGAAGGAAGAGTCTGTCAAGAAAGAT 2274

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   2275 TCTGTTGCTGCCGGAAAGCCAAAGGAGAAGGGGAAAATAAAAGTCATTAAGAAGGAAGGC 2334

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   2335 AAGGCCGCAGAGGCTGTCGCTGCAGCTGTCGGCACTGGAGCCACCACAGCAGCTGTCATG 2394

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   2395 GCGGCAGCTGGAATAGCAGCC---------ATTGGCCCTGCCAAAGAACTCGAAGCTGAG 2445

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGln---LeuThrAspLeuSerProGlu 471
Db   2446 AGGTCCCTTATGTCATCTCCTGAGGATCTAACCAAGGACTTTGAAGAGTTAAAGGCTGAA 2505

Qy    472 GluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSer 491
Db   2506 GAGGTCGATGTAACAAAG---------GACATCAAGCCTCAGCTGGAGCTAATCGAAGAC 2556

Qy    492 GlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeu 511
Db   2557 GAAGAGAAACTGAAG---GAAACTGAGCCAGTCGAAGCCTACGTCATCCAGAAGGAGAGA 2613

Qy    512 LysLysLeuSerGlyLysLysGlnLysGlyLysArgGly--------------------- 524
Db   2614 GAAGTCACCAAAGGTCCTGCCGAGTCCCCTGATGAGGGAATCACTACCACTGAAGGGGAG 2673

Qy    525 GlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSer 544
Db   2674 GGCGAATGTGAACAGACACCTGAGGAGCTGGAGCCCGTCGAGAAGCAGGGAGTAGAC--- 2730

Qy    545 AlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThr 564
Db   2731 ---GACATTGAAAAAATTTGAAGATGAAGGAGCCGGTTTTGAAGAATCTTCAGAGACT--- 2784

Qy    565 CysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThr 584
Db   2785 ------------GGAGACTATGAAGAGAAGGCAGAAACTGAGGAGGCTGAGGAGCCAGAA 2832

Qy    585 SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetVal 604
Db   2833 GAGGATGGGGAGGAACAC---------------------------------------GTA 2853

Qy    605 ThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeuGlu--- 623
Db   2854 TGTGTGAGCGCCTCCAAGCACAGCCCCACTGAGGATGAGGAAAGTGCCAAGGCGGAGGCT 2913

Qy    624 ------------LysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGlu 639
Db   2914 GATGCATACATCAGGGAGAAGAGGGAGTCTGTGGCCAGTGGGGATGACCGAGCCGAAGAA 2973

Qy    640 MetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArg 659
Db   2974 GACATGGATGAGGCCATTGAGAAAGGAGAGGCTGAACAATCTGAAGAGGAGGCTGATGAG 3033

Qy    660 ValAspThrSer------------------------------ValSerTrpGluAlaLeu 669
Db   3034 GAGGACAAAGCTGAAGATGCCAGAGAGGAGGAATATGAGCCGGAAAAAATGGAAGCTGAA 3093

Qy    670 IleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGly 689
Db   3094 GACTATGTGATGGCTGTGGTCGACAAGGCTGCAGAGGCTGGTGGTGCCGAGGAGCAGTAT 3153

Qy    690 Gly---------------ProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAla 704
Db   3154 GGATTCCTCACCACACCAACCAAGCAACTAGGAGCCCAGTCTCCTGGCCGAGAACCTGCA 3213

Qy    705 SerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGln 724
Db   3214 TCTTCAATTCATGAT---------GAGACTTTACCTGGAGGCTCAGAGAGCGAGGCCACC 3264

Qy    725 AlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSer 744
Db   3265 GCTTCTGATGAGGAGAATCGAGAA------GACCAGCCTGAGGAATTCACTGCCACCTCT 3318

Qy    745 ThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGlu 764
Db   3319 GGCTAC------------------------------------ACTCAGTCTACTATTGAG 3342

Qy    765 GluLysAlaGluAspSerSerValGluGlnLeuSerThr--------------------- 777
Db   3343 ATATCCAGTGAGCCCACCCCCATGGATGAGATGTCTACCCCTCGAGACGTGATGAGTGAT 3402

Qy    778 ------------GluIleGluProSerArgGluGluSerTrpValSerIleLysLys--- 792
Db   3403 GAGACCAACAATGAAGAGACGGAGTCCCCTTCTCAGGAATTCGTAAATATCACCAAATAT 3462

Qy    792 ------------------------------------------------------------ 792
Db   3463 GAATCTTCATTGTATTCTCAGGAATACTCTAAACCTGCTGATGTTACACCGCTCAACGGA 3522
```

```
Qy    793 PheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGlu 812
Db   3523 TTTTCTGAAGGATCAAAAACAGATGCCACTGATGGCAAGGATTACAATGCTTCAGCCTCT 3582
Qy    813 AspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValValProLeu 832
Db   3583 ACCATATCACCACCC--------------------------------------------- 3597
Qy    833 SerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGluLeuPro 852
Db   3598 ---------TCTTCCATGGAGGAAGACAAATTCAGCAGATCTGCT--------------- 3633
Qy    853 GlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrVal 872
Db   3634 ------TTACGTGATGCTTACTGCTCTGAAGTGAAAGCCAGCACC--------------- 3672
Qy    873 SerValAlaValIleAspGlyThrArgAlaValThrSerValGlu---------GluArg 889
Db   3673 ACTTTGGACATCAAAGATAGCATCTCAGCTGTTTCAAGTGAAAAGGTCAGCCCATCGAAG 3732
Qy    890 SerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr---AlaGlyGlu 908
Db   3733 AGCCCGTCC---CTGAGTCCATCTCCACCATCACCCTTAGAAAAGACCCCCCTGGGTGAA 3789
Qy    909 ---------------AlaMetProProValGlu---GluValThrGluLysAspIleIle 922
Db   3790 CGTAGTGTGAACTTCTCTCTGACGCCCAATGAGATTAAAGTCTCTGCAGAGGCAGAAGTA 3849
Qy    923 AlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAsp 942
Db   3850 GCCCCGGTGTCTCCTGAGGTGACCCAAGAAGTAGTTGAAGAACATTGTGCTAGTCCTGAG 3909
Qy    943 MetValThrSerGluValAspPheThrSerGluAlaValThrAlaThr------------ 958
Db   3910 GACAAGACTCTGGAAGTGGTGTCACCATCTCAGTCCGTGACTGGCAGTGCTGGTCACACA 3969
Qy    959 ------------------------GluThrSerGluAlaLeuArgThrGluGluVal--- 969
Db   3970 CCTTACTATCAATCTCCTACTGACGAGAAATCCAGTCATCTCCCTACAGAAGTCATTGAA 4029
Qy    969 ------------------------------------------------------------ 969
Db   4030 AAACCACCAGCAGTTCCAGTGAGTTTTGAATTCAGTGATGCCAAAGATGAGAATGAAAGG 4089
Qy    969 ------------------------------------------------------------ 969
Db   4090 GCTTCAGTAAGCCCCATGGATGAGCCCGTGCCTGACTCAGAGTCTCCTATTGAAAAAGTT 4149
Qy    969 ------------------------------------------------------------ 969
Db   4150 TTGTCTCCTTTACGCAGCCCGCCCCTCATTGGATCCGAGTCTGCTTATGAAAGTTTTCTA 4209
Qy    970 ---------------------------------------------ThrGluAlaSerGly 974
Db   4210 AGTGCTGATGACAAGGCTTCTGGCAGAGGTGCCGAAAGTCCTTTTGAAGAAAAGAGTGGA 4269
Qy    975 AlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSer--------- 991
Db   4270 AAACAAGGCTCTCCAGACCAAGTAAGTCCAGTTTCTGAAATGACTTCTACTAGTCTTTAC 4329
Qy    992 ProAspThrThrGlu------------GluAlaThrProValGlnGluValGluSerGly 1007
Db   4330 CAAGACAAACAGGAAGGGAAAAGCACAGACTTTGCACCAATAAAAGAAGACTTTGGCCAA 4389
Qy   1008 ValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAsp 1027
Db   4390 GAAAAGAAAACTGATGATGTTGAAGCCATGAGTTCTCAACCAGCACTGGCTCTGGATGAA 4449
Qy   1028 Lys--------------------------------------------------------- 1028
Db   4450 AGGAAATTAGGAGATGTTTCTCCCACACAAATAGATGTCAGTCAGTTTGGATCTTTTAAA 4509
Qy   1029 ---------------------------------ValLysGluGluSerGlnValProAla 1037
Db   4510 GAAGACACTAAGATGTCCATTTCTGAAGGTACTGTCTCAGACAAGTCAGCTACTCCTGTT 4569
Qy   1038 ThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   4570 GATGAGGGCGTAGCAGAAGACACGTACTCTCATATGGAGGGTGTGGCCTCAGTGTCCACA 4629
Qy   1058 AspSerGluValLeuAlaSerGluLysGlu---------LysAspValMetProLysGly 1074
Db   4630 GCCTCAGTGGCTACGAGCTCATTTCCAGAGCCAACAACAGATGATGTGTCTCCATCTCTG 4689
Qy   1075 ProValGlnGluAlaGlyAlaGluHis--------------------------------- 1083
Db   4690 CATGCT---GAGGTTGGCTCCCCACATTCCACAGAAGTAGATGACTCCCTTTCAGTGTCT 4746
Qy   1084 LeuAlaGlnGlySerGluThrGlyGlnAlaThr--------------------------- 1094
Db   4747 GTTGTGCAAACACCTACCACATTCCAGGAAACAGAAATGTCTCCATCTAAAGAAGAATGC 4806
Qy   1095 ------ProGluSerLeuGluValProGluVal------------------------Thr 1104
Db   4807 CCAAGACCGATGTCAATTTCTCCACCAGATTTCTCCCCTAAAACTGCAAAGTCCAGGACA 4866
Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeu---------GlnGlnLeu 1121
Db   4867 CCCGTTCAAGATCACAGATCTGAACAGTCCTCAATGTCTATTGAATTTGGCCAAGAATCT 4926
Qy   1122 MetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThr------ 1139
Db   4927 CCTGAGCAATCCCTTGCTATGGACTTCAGTCGACAGTCTCCAGATCACCCTACAGTGGGT 4986
Qy   1140 ------------------------AsnGlySerThrProLeu------AlaAspSerAsp 1149
Db   4987 GCAGGCGTGCTTCACATCACTGAAAATGGGCCAACTGAAGTGGACTACAGTCCTTCTGAC 5046
Qy   1150 ThrAlaAspGlyThr------------------GlnGlnAspGluThrIleAspSerGln 1163
Db   5047 ATGCAGGACTCCAGTTTATCACATAAGATACCACCTATGGAGGAGCCGTCCTACACCCAA 5106
Qy   1164 AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThr 1183
Db   5107 GATAATGATCTTTCTGAGCTCATCTCAGTATCTCAGGTA---GAGGCCTCCCCGTCCACC 5163
Qy   1184 AlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGly 1203
Db   5164 TCTTCTGCTCATACCCCTTCTCAGATCGCTTCT------CCTCTCCAA------------ 5205
Qy   1204 GluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaVal 1223
Db   5206 GAAGATACTCTATCCGATGTTGCTCCTCCCAGAGATATGTCCTTATATGCCTCACTCACC 5265
Qy   1224 ProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGlu 1243
Db   5266 ------------------------------TCTGAAAAAGTGCAAAGTCTGGAAGGAGAG 5295
Qy   1244 LysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAla 1263
Db   5296 AAGCTCTCTCCAAAATCTGATATCTCTCCACTCACCCCACGAGAGTCCTCTCCTTATAT 5355
Qy   1264 AspValThrTyrAspSerGluValMetGlyVal------------AlaGlyCysGlnGlu 1279
Db   5356 TCACCTACTTTTTCAGATTCTACCTCTGCAGTCAAAGAGAAAACAGCAACTTGCCACAGT 5415
Qy   1280 LysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspVal 1299
Db   5416 TCCTCTTCTCCACCAATAGATGCAGCATCCGCAGAGCCCTATGGCTTCCGTGCCTCAGTG 5475
Qy   1299 ------------------------------------------------------------ 1299
Db   5476 TTATTCGATACAATGCAACACCATCTAGCCTTGAATAGAGATTTGTCCACACCTGGCCTG 5535
Qy   1300 GluLysGluLysArgGluThrLysProGluGlnValSer---------GluGluGlyGlu 1316
```

```
Db   5536 GAGAAGGACAGTGGAGGGAAGACACCTGGTGACTTTAGCTATGCCTATCAAAAGCCTGAG 5595

Qy   1317 GlnGluThrAlaAlaProGluHisGluGly------------ThrTyrGlyLysProVal 1332
             :::   |||   :::|||:::   |||               :::|||  |||
Db   5596 GAAACAACCAGGTCCCCAGATGAAGAAGATTATGACTATGAGTCTTATGAGAAGACCACC 5655

Qy   1333 LeuThrLeuAspMet---------------------------------------ProSer 1339
             |||   |||:::
Db   5656 CGGACCTCAGATGTGGGTGGCTATTACTATGAGAAGATAGAGAGAACCACAAAATCTCCA 5715

Qy   1340 SerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAsp 1359
          |||:::   |||   :::    ::::::|||   :::        |||:::
Db   5716 AGTGACAGTGGCTACTCCTATGAGACCATTGGGAAAACTACCAAGACCCCTGAAGATGGT 5775

Qy   1360 LysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAla 1379
                                                                  :::
Db   5776 ------------------------------------------------GACTATTCCTAT 5787

Qy   1380 GluAlaValGluLysValIleGluThrVal------------------------ValIle 1391
          |||   :::||||||         |||                           :::
Db   5788 GAAATTATTGAGAAGACCACACGGACCCCTGAAGAGGGTGGGTACTCATATGACATAAGT 5847

Qy   1392 SerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLysSer 1411
             :::|||          |||||     |||                 |||:::|||
Db   5848 GAAAAGACCACCAGCCCCCCGAAGTGAGTGGTTACAGCTATGAAAAGACTGAGAGGTCT 5907

Qy   1412 ---------------------------------------SerAlaThrGlyGlyHisTrp 1418
                                                 |||      |||||||||
Db   5908 AGAAGGCTTCTGGATGACATCAGCAATGGCTATGATGACTCTGAGGATGGTGGCCAC--- 5964

Qy   1419 ThrLeu------------------GlnHisAlaGluAspThrValProLeuGlyProGlu 1432
          ||||||                                  :::   :::      |||||
Db   5965 ACACTTGGGGACCCCAGCTACTCTTATGAAACCACTGAGAAAATTACCAGTTTCCCTGAG 6024

Qy   1433 SerGlnAlaGluSerIleProIleIle------------------------------- 1441
          |||:::      |||
Db   6025 TCTGAAGGTTATTCCTATGAGACATCTACAAAGACAACACGAACCCCTGATACTTCCACA 6084

Qy   1442 ------------------------ValThrProAlaProGluSerThrLeuHisProAsp 1453
                                  :::|||     |||:::::::::   :::
Db   6085 TACTGTTACGAGACTGCAGAGAAAATCACTAGAACCCCTCAGGCATCCACATATTCCTAC 6144

Qy   1454 LeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysProAsp 1473
                :::::        :::       ::::::|||      |||    :::   |||
Db   6145 GAGACTTCAGACCTATGCTACACTGCAGAAAAGAAGTCCCCCTCAGAAGCCCGTCAGGAT 6204

Qy   1474 AlaGly---------------------------------ProAspAlaAspGlyLysGlu 1482
                                                   |||       :::
Db   6205 GTCGATTTATGCCTCGTGTCCTCTTGTGAATACAAGCACCCCAAGACAGAGCTTTCACCC 6264

Qy   1483 Ser-ThrAlaIleGluLysValLeu---------------------------------- 1490
          ||| :::    |||       :::|||
Db   6265 TCTTTCATTAATCCCAATCCTCTTGAGTGGTTTGCCAGTGAAGAACCCACTGAAGAATCT 6324

Qy   1491 ------------------LysAlaGluProGluIleLeuGluLeuGluSerLysSerAs 1504
                            :::    ||||||      |||:::    ||||||
Db   6325 GAAAAGCCCCTCACTCAATCAGGGGGAGCCCCACCGCCTCCAGGAGGAAAGCAACAGGGC 6384

Qy   1504 nLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThr-- 1523
                   |||:::        :::    |||:::|||    :::       ::::::|||
Db   6385 CGACAGTGTGATGAAACCCCTCCC-ACCTCAGTCAGCGAGTCAGCCCCATCCCAGACCGA 6443

Qy   1524 ---------AlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysAr 1540
                             |||||||||          |||    |||
Db   6444 CTCTGATGTTCCCCCGGAGACTGAAGAGTGCCCCTCCATCACG------GCCGATGCCAA 6497

Qy   1540 gLeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHi 1560
            :::||||||   :::   :::                  |||         |||||
Db   6498 TATCGACTCTGAAGACGAGTCGGAAACCATCCCCACAGACAAAACTGTCACGTACAAACA 6557

Qy   1560 s------------------ProVal------------ProGlnProArgGlu-------- 1567
          |                  ||||||             |||   ||||||
Db   6558 CATGGACCCACCTCCAGCTCCCGTGCAAGACCGCAGCCCTTCGCCACGCCACCCTGATGT 6617

Qy   1568 ----------AspLeuGlnValLeuThrValLeuGlu-AlaTrpAlaGln---------- 1580
                    |||    :::   |||    :::    :::    ||||||:::
Db   6618 GTCCATGGTGGACCCAGAGGCCTTGGCCATTGAGCAGAACCTGGGCAAAGCTCTAAAGAA 6677

Qy   1581 --------ProArgLysCysLeuProArgLeuGlnLeuLysAlaPro 1593
                  |||:::      |||:::   |||::::::    |||
Db   6678 AGATCTGAAAGAGAAGACCAAAACCAAAAAGCCAGGTACAAAGACCA 6724
```

RESULT 8

US-10-203-138A-6112

```
; Sequence 6112, Application US/10203138A
; GENERAL INFORMATION:
;  APPLICANT: Molecular Dynamics, Inc.
;  APPLICANT:  Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  ANALYSIS OF GENE EXPRESSION IN BT 474
;  FILE REFERENCE: PB 0004 WO 8
;  CURRENT APPLICATION NUMBER: US/10/203,138A
;  CURRENT FILING DATE:  2002-08-02
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 04 February 2000 (04.02.00)
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 26 May 2000 (26.05.00)
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 03 August 2000 (03.08.00)
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 03 October 2000 (03.10.00)
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 27 September 2000 (27.09.00)
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 21 September 2000 (21.09.00)
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 30 June 2000 (30.06.00)
;  NUMBER OF SEQ ID NOS: 15438
;  SOFTWARE: Molecular Dynamics Sequence Listing Engine
; SEQ ID NO 6112
;   LENGTH: 7997
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL034555.2
;   FEATURE:
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  12
;   FEATURE:
;   OTHER INFORMATION: EST_HUMAN HIT: AU117052.1, EVALUE 0.00e+00
;   FEATURE:
;   OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUE 4.00e-10
;   FEATURE:
;   OTHER INFORMATION: NT HIT: gi11427711, EVALUE 0.00e+00
US-10-203-138A-6112
```

Alignment Scores:

Pred. No.:	4.03e-10	Length:	7997
Score:	431.50	Matches:	373
Percent Similarity:	34.27%	Conservative:	271
Best Local Similarity:	19.85%	Mismatches:	759
Query Match:	5.34%	Indels:	476
DB:	6	Gaps:	83

US-09-902-432-4 (1-1596) x US-10-203-138A-6112 (1-7997)

```
Qy     31 ProAlaAlaGluAlaSerGlyAlaAlaGly-AspProAlaAspAlaAspProAla----- 48
             |||      |||   ||||||    :::      :::||||||    :::|||   :::
Db    886 CCTGGACCAGAAACTTCAGGTCTCTCAGACGGAGCCTGCAAAATCTGACTTGTCTAAACT 945

Qy     49 -------------ThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAsnGlyVa 64
                       |||:::|||::::::   |||        |||   |||       ||
Db    946 GGAATCAGTTAGAATGAAAGTACCAAAGGAAAAGGGGCTTTCAAGCCATGTTGAAGTGGT 1005
```

```
Qy     64 lAlaGluGlnGlyAspVal------------HisValGlnGluGluAsnGlnGluGlyGl 80
Db   1006 GGAGAAGGAAGGCAGGCTTAAAGCCAGGAAGCACCTCAAGCCTGAG---CAGCCTGCAGA 1062

Qy     80 nGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLy 100
Db   1063 TGGGGTAAGTGCTGTGGAT----------------------------------------- 1081

Qy    100 sAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGl 120
Db   1082 -------CTGGAGAAGCTGGAAGCCAGGAAAAGGCGCTTTGCAGATTCCAATTTAAAAGC 1134

Qy    120 yGlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsn---AsnValGl 139
Db   1135 AGAAAAGCAAAAACCAGAGGTCAAGAAAAGCAGTCCAGAGATGGAGGATGCTCGCGTGCT 1194

Qy    139 uGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPhe-- 158
Db   1195 TTCAAAAAAGCAGCCT------------------GACGTGTCCTCTAGAGAGGTCATTCT 1236

Qy    159 -------------------------LysPheValGlyPheLysPheThrValLysLysAs 170
Db   1237 GCTGAGGGAAGGAGAGGCTGAAAGAAAGCCTGTGAGGAAAGAAATTCTTAAAAGAGAATC 1296

Qy    170 pLysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGl 190
Db   1297 TAAAAAAATCAAACTGGACAGACTTAATACTGTTGCCAGCCCCAAAGACTGTCAGGAGCT 1356

Qy    190 yAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGl 210
Db   1357 TGCCAGTATTTCTGTTGGGTCTGGCTCAAGGCCCAGCTCAGACCTACAAGCAAGACTGGG 1416

Qy    210 yGluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLe 230
Db   1417 AGAACTAGCAGGT------------------GAATCTGTGGAAAATCAAGAA-------- 1450

Qy    230 uLysGlnGluGlnSerSerThrGluIlePro---------------------LeuGlnAl 243
Db   1451 -------GTCCAATCAAAAAAGCCCATTCCCTCAAAACCACAGCTCAAACAGCTGCAGGT 1503

Qy    243 aGluSerAspGlnAlaAlaGluGluGlu------AlaLysAspGluGlyGluGluLysGl 261
Db   1504 ATTAGATGATCAAGGACCAGAGAGAGAAGACGTTAGGAAAAACTATTGCAGTCTTCGTGA 1563

Qy    261 nGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr-- 280
Db   1564 TGAAACACCTGAACGTAAATCAGGCCAAGAGAAATCACATTCAGTAAATACTGAAGAAAA 1623

Qy    281 ----------------------ThrSerSerPheLysLysPhePheThrHisGlyTrpAl 293
Db   1624 AATTGGCATTGACATCGATCACACGCAGAGTTACCGAAAACAAATGGAACAGAGTCGTAG 1683

Qy    293 aGlyTrpArgLysLysThrSerPheLysLysSerLys---------------GluAspAs 308
Db   1684 GAAACAGCAGATGGAAATGGAAATAGCCAAGTCTGAGAAGTTTGGCAGTCCTAAAAAAGA 1743

Qy    308 pLeuGluThrAlaGluLysArg-------------------------------------- 315
Db   1744 TGTAGATGAATATGAAAGACGTAGCCTCGTTCACGAGGTAGGCAAACCCCCTCAAGATGT 1803

Qy    316 ----------------------LysGluGlnGluAlaGluLysValAsp----------- 324
Db   1804 CACTGATGACTCTCCTCCTAGCAAAAAGAAAAGGATGGATCATGTCGATTTTGATATCTG 1863

Qy    325 ----------GluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluProAl 341
Db   1864 CACCAAGCGAGAACGGAATTACAGAAGTTCACGCCAAATCAGCGAAGATTCTGAAAGGAC 1923

Qy    341 aGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGl 361
Db   1924 TGGTGGTTCTCCCAGTGTCCGACATGGTTCCTTCCATGAAGAT---------------GA 1968

Qy    361 uAspGlnValGlyAspLeuGluAlaSerSer---------------------GluGluLy 374
Db   1969 GGATCCCATAGGCTCCCCTAGGCTACTGTCAGTAAAAGGGTCTCCTAAAGTAGATGAAAA 2028

Qy    374 sCysAlaPro------------------------------LeuAlaThrGluValPheAs 384
Db   2029 AGTCCTCCCCTATTCTAACATAACAGTCAGGGAAGAGTCTTTAAAATTTAATCCTTATGA 2088

Qy    384 pGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSerThrValGluLy 404
Db   2089 TTCTAGCAGGAGAGAACAGATGGCAGATATGGCCAAAATAAAACTATCTGTCTTGAATTC 2148

Qy    404 sThrGluGluGluGlnGlyGlyGlyGlyGluAlaGlu---------GlyGlyValValVa 421
Db   2149 TGAAGATGAACTAAATCGTTGGGACTCTCAGATGAAACAGGATGCTGGCAGATTTGATGT 2208

Qy    421 lGluGlyThrGlySerLeuProProGluLysLeu------------------------Al 434
Db   2209 GAGTTTCCCAAACAGCATAATTAAGAGAGATAGCCTTCGAAAAAGGTCTGTACGAGATCT 2268

Qy    434 aGluProGlnGluValProGlnGluAlaGluProAlaGluGlu----------------- 448
Db   2269 GGAACCTGGTGAGGTGCCTTCTGATTCTGACGAAGATGGTGAACACAAATCCCACTCACC 2328

Qy    449 ----------------------------------------LeuMetLysSerArgGluMe 455
Db   2329 CAGAGCCTCTGCATTATATGAAAGTTCTCGATTGTCTTTTTTATTGAGGGACAGAGAAGA 2388

Qy    455 tCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGlu----------- 471
Db   2389 CAAGCTACGTGAGCGAGATGAAAGACTCTCTAGTTCTTTAGAAAGGAACAAATTTTACTC 2448

Qy    472 ----------GluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMe 488
Db   2449 TTTTGCATTGGATAAGACAATCACACCAGACACTAAAGCTTTGCTTGAAAGAGCTAAATC 2508

Qy    488 tLeuSerSer-------------------------------------------------- 491
Db   2509 CCTCTCTTCATCTCGTGAAGAAAATTGGTCTTTTCTTGATTGGGACTCCCGATTTGCAAA 2568

Qy    492 -------------GlnGluArgIleLysValGlnGlySerPro----------------- 501
Db   2569 TTTTCGAAACAACAAAGATAAAGAAAAGGTTGACTCTGCTCCAAGACCTATTCCATCCTG 2628

Qy    502 -------LeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGl 519
Db   2629 GTACATGAAAAAGAAGAAAATTAGGACTGATTCAGAAGGGAAAATGGATGATAAGAAAGA 2688

Qy    519 nLysGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisTh 539
Db   2689 GGACCATAAA---------------GAAGAAGAGCAAGAGAGGCAGGAATTGTTTGCTTC 2733

Qy    539 rGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGl 559
Db   2734 TCGTTTTTTACACAGCTCAATCTTTGAACAAGATTCCAAGCGATTGCAGCATCTAGAGAG 2793

Qy    559 uProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAl 579
Db   2794 AAAAGAGGAAGATTCTGACTTCATTTCTGGTAGGATCTATGGGAAGCAGACATCTGAGGG 2853

Qy    579 aGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSe 599
Db   2854 AGCAAACAGCACAACT---------GATTCCATTCAAGAACCAGTAGTTCTGTTCCATAG 2904

Qy    599 r---PheLysLysMetVal------------ThrProLysLysArgValArgArgProSe 614
Db   2905 CAGATTTATGGAGCTCACACGGATGCAACAGAAAGAAAAAGAAAAAGACCAGAAACCCAA 2964

Qy    614 rGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAs 634
Db   2965 AGAGGTTGAGAAACAGGAAGATACAGAGAATCATCCCAAGACCCCAGAATCTGCTCCTGA 3024

Qy    634 pSerThrValSerGluMetGln---------------------AspGluValLysThrVa 647
```

```
Db   3025 GAATAAAGATTCAGAACTGAAAACTCCACCTTCCGTTGGGCCTCCAAGTGTCACAGTCGT 3084

Qy    647 lGlyGluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGl 667

Db   3085 AACTCTAGAATCAGCCCCATCAGCACTAGAGAAGACCACTGGTGACAAAACGGTAGAGGC 3144

Qy    667 uAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAs 687

Db   3145 GCCTTTGGTA---------ACAGAAGAGAAGACTGTGGAGCCAGCTACCGTCTCAGAAGA 3195

Qy    687 p---------------GluGlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGl 702

Db   3196 AGCAAAGCCTGCATCTGAACCTGCTCCTGCCCCTGTG------------GAACAGCTGGA 3243

Qy    702 uGlu------------------AlaSerLysAspLysGluAlaGlyThrAspAlaValPr 716

Db   3244 ACAAGTAGACCTGCCCCCAGGAGCAGACCCCGATAAAGAAGCTGCC------ATGATGCC 3297

Qy    716 oAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluPro-------- 733

Db   3298 TGCGGGTGTTGAGGAA---------------GGTTCATCAGGTGACCAGCCGCCTTATCT 3342

Qy    734 ----AlaGlySerProSerGluGlyGluGlyValSerThrTrpGlu-------------- 747

Db   3343 GGATGCCAAGCCTCCAACTCCCGGGGCCTCGTTTTCCCAGGCAGAGAGCAACGTAGATCC 3402

Qy    748 ----------SerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGl 764

Db   3403 AGAGCCTGACAGTACCCAGCCACTTTCAAAACCAGCTCAGAAGTCTGAGGAAGCCAATGA 3462

Qy    764 uGluLysAlaGlu------AspSerSerValGlu------------------------- 773

Db   3463 GCCAAAGGCCGAAAAGCCAGACGCCACTGCAGATGCTGAGCCTGATGCAAACCAGAAAGC 3522

Qy    774 -GlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPh 793

Db   3523 CGAAGCTGCTCCTGAGTCTCAGCCCCCAGCTTCTGAAGAT---TTAGAGGTTGATCCTCC 3579

Qy    793 eIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAs 813

Db   3580 AGTTGCTGCAAAGGATAAAAAGCCAAACAAAAGCAAGCGTTCAAAGACCCCTGTTCAG-- 3637

Qy    813 pSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValValProLeuSe 833

Db   3638 -GCAGCTGCAGTGAGTATCGTGGAG---------------AAGCCCGTCACAAGGAAGAG 3681

Qy    833 rGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThr------GluLe 851

Db   3682 TGAG------AGGATAGACCGGGAAAAACTCAAGCGGTCCAATTCTCCTCGGGGAGAAGC 3735

Qy    851 uProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThr----------- 867

Db   3736 ACAGAAGCTTTTGGAATTGAAGATGGAGGCAGAGAAGATTACAAGGACTGCTTCTAAAAA 3795

Qy    868 -------------LeuValHisThrValSerValAlaValIleAspGlyThrArgAla-- 882

Db   3796 CTCTGCTGCAGACCTTGAACATCCCGAACCAAGTTTGCCTCTCAGCCGAACAAGGCGCCG 3855

Qy    883 ----ValThrSerVal---------------------GluGluArgSerProSerTrpIl 894

Db   3856 GAATGTAAGGAGCGTCTATGCAACCATGGGTGACCATGAAAACCGCTCTCCT-------- 3907

Qy    894 eSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGl 914

Db   3908 ----------GTCAAAGAGCCCGTTGAGCAA---------------------------CC 3930

Qy    914 uGluValThrGluLysAspIleIleAlaGlu------GluThrProValLeuThrGlnTh 932

Db   3931 AAGAGTGACCAGAAAGAGATTGGAGCGAGAGCTTCAGGAGGCTGCAGCGGTTCCCACCAC 3990

Qy    932 r---------------LeuProGluGlyLysAspAlaHisAspAspMetValThrSerGl 947

Db   3991 CCCTCGGAGGGGAAGGCCTCCAAAGACACGCCGGCGAGCCGATGAAGAGGAGGAGAACGA 4050

Qy    947 uValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGl 967

Db   4051 GGCCAAGGAACCTGCAGAAACACTCAAGCCACCTGAGGGATGGCGGTCGCCAAGGTCCCA 4110

Qy    967 uGluValThrGluAlaSerGlyAlaGlu------------GluThrThrAspMetValSe 983

Db   4111 GAAAACTGCAGCTGGTGGTGGACCCCAAGGGAAAAAGGGAAAAAATGAACCGAAGGTGGA 4170

Qy    983 rAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGl 1003

Db   4171 TGCTACA---------------CGTCCTGAGGCCACCACTGAGGTGGGCCCCCAAATAGG 4215

Qy   1003 uValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGl 1023

Db   4216 CGTGAAAGAGAGCTCCATGGAACCCAAGGCTGCTGAGGAGGAGGCAGGGAGTGAACAGAA 4275

Qy   1023 nAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnAr 1043

Db   4276 ACGTGACAGAAAAGATGCTGGCACAGACAAAAACCCCCCTGAAACCGCCCCTGTTGAA-- 4333

Qy   1043 gThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSer----------- 1059

Db   4334 -------------GTTGTAGAGAAAAAACCGGCCCCTGAAAAAAACTCCAAATCAAAGAG 4380

Qy   1060 ------------------------------GluValLeuAlaSerGluLysGluLysAs 1069

Db   4381 AGGAAGATCTCGAAACTCCAGGTTAGCAGTGGACAAATCTGCAAGTCTGAAAAATGTGGA 4440

Qy   1069 p------ValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGl 1087

Db   4441 TGCTGCTGTCAGTCCCAGGGGGGCTGCAGCACAGGCAGGGGAGAGG-------------- 4486

Qy   1087 ySerGluThrGly------GlnAlaThrProGluSerLeuGluValProGluValThrAl 1105

Db   4487 ----GAATCTGGGGTGGTGGCAGTCTCCCCTGAGAAAAGTGAGAGTCCCCAAAAGGAGGA 4542

Qy   1105 aAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnAl 1125

Db   4543 TGGTTTATCA-----------------------------TCCCAGTTGAAAAGTGATCC 4572

Qy   1125 aValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLe 1145

Db   4573 AGTTGATCCAGACAAGGAACCAGAGAAAGAAGACGTGTCTGCCTCTGGGCCGTCCCCAGA 4632

Qy   1145 uAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSer-------- 1162

Db   4633 AGCC---ACCCAGTTAGCCAAGCAGATGGAGCTGGAGCAGGCCGTGGAACACATCGCAAA 4689

Qy   1163 ----GlnAspSerLysAlaThrAlaAlaValArgGlnSer-------------------- 1174

Db   4690 GCTCGCTGAGGCCTCTGCCTCTGCTGCCTATAAGGCAGATGCACCAGAGGGCCTTGCCCC 4749

Qy   1175 -------------------------GlnValThrGluGluGluAlaAlaThrAlaGlnLy 1186

Db   4750 AGAGGACAGGGACAAGCCTGCACACCAAGCAAGTGAAACAGAGCTGGCTGCGGCCATCGG 4809

Qy   1186 s---------GluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHis-- 1202

Db   4810 CTCCATCATCAATGACATTTCTGGGGAGCCAGAAAACTTCCCAGCACCTCCACCTTATCC 4869

Qy   1203 -GlyGluGlu--------------------ProGlyArgAspValLeuGluProThrGl 1215

Db   4870 TGGAGAATCCCAGACAGATCTGCAACCCCCCGCAGGTGCACAGGCGCTGCAGCCTTCTGA 4929

Qy   1215 nGlnGluLeu---ThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGl 1234

Db   4930 GGAAGGAATGGAGACAGATGAGGCTGTATCTGGCATCCTGGAAACTGAGGCTGCTACAGA 4989

Qy   1234 uGlyGluValAspTrpLeuAspGlyGluLysValLysGluGluGlnGluValPheValHi 1254

Db   4990 ATCTTCTAGGCCTCCAGTCAATGCTCCTGACCCCTCA---------------------- 5026
```

```
Qy   1254 sSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyVa 1274
           :::||||||        |||  |||                              ||
Db   5027 -GCCGGCCCAACAGATACCAAGGAAGCCAGAGGAAATAGCAGTGAAACCTCACACTCAGT 5085

Qy   1274 lAlaGlyCysGlnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGl 1294
          |          :::     :::    ||||||     :::        ::::::|||
Db   5086 GCCAGAAGCCAAAGGGTCTAAAGAAGTGGAAGTCACTCTTGTTCGGAAAGACAAAGGGCG 5145

Qy   1294 uMetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGl 1314
             :::|||        :::    ||||||  |||      :::   |||:::
Db   5146 CCAGAAAACAACCCGATCACGCCGCAAGCGAAACACAAACAAGAAAGTGGTGGCT----- 5200

Qy   1314 uGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuTh 1334
                      |||:::      ||||||              |||
Db   5201 ----CCTGTAGAGAGCCATGTCCCTGAATCCAACCAAGCTCAAGGT------------- 5242

Qy   1334 rLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSe 1354
             :::    |||::::::      |||      :::                |||
Db   5243 ----GAGAGTCCTGCTGCAAATGAGGGGACAACAGTA-----------CAGCACCCCGA 5286

Qy   1354 rLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrTh 1374
             |||    ::::::|||                    |||    :::       :::||
Db   5287 AGCCCCACAGGAAGAAAAG-----------------CAGAGTGAGAAACCCCATTCCAC 5328

Qy   1374 rValThrGlnThrAla--------------------GluAlaValGluLysValIleGl 1387
          |      |||:::                       :::    |||          ::
Db   5329 TCCTCCTCAGTCATGTACTTCTGACCTAAGCAAGATTCCCTCCACAGAGAATTCGTCCCA 5388

Qy   1387 uThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeu----- 1405
          :   :::   :::   |||          :::         |||        |||
Db   5389 AGAAATCAGTGTTGAGGAAAGGACTCCAACCAAAGCATCTGTGCCCCCAGACCTTCCCCC 5448

Qy   1406 -------LeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAl 1423
                      ||||||                           :::    :::
Db   5449 ACCTCCCCAGCCAGCACCGGTGGATGAGGAGCCTCAAGCCAGGTTCAGGGTGCATTCCAT 5508

Qy   1423 aGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValTh 1443
             ::::::     |||:::    |||   |||          |||||||||||
Db   5509 CATTGAAAGTGACCCGGTGACCCCACCCAGCGAT---CCAAGCATCCCCATACCCACACT 5565

Qy   1443 rProAla-----------------------------------------ProGluSe 1448
            |||:::                                          |||
Db   5566 GCCTTCTGTAACTGCAGCAAAGCTCTCACCTCCTGTCGCCTCTGGGGGGATCCCACACCA 5625

Qy   1448 rThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGl 1468
           :::          :::          |||:::       :::     |||::::::
Db   5626 GAGCCCCCCTACTAAGGTGACAGAGTGGATCACAAGGCAGGAGGAGCCACGGGCTCAGTC 5685

Qy   1468 uGluAspLysProAspAlaGlyProAspAlaAspGlyLysGlu-----------SerTh 1484
                      |||        ||||||           :::          |||::
Db   5686 TACTCCATCTCCAGCTCTTCCCCCAGACACAAAGGCCTCTGATGTTGACACCAGCTCCAG 5745

Qy   1484 rAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAs 1504
          :   :::   |||:::|||        :::|||:::   :::          |||   :::::
Db   5746 CACCCTGAGGAAGATTCTC---ATGGACCCCAAGTATGTGTCTGCCACAAGTGTCACTTC 5802

Qy   1504 nLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAl 1524
          :        |||         |||         |||                     ||
Db   5803 CACAAGTGTCACCACAGCCATTGCAGAGCCTGTC-----------------AGTGCTGC 5844

Qy   1524 aProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerAr 1544
          ||||      |||                   :::    |||       :::|||||||::
Db   5845 CCCTTGCCTACAT-----------------GAGGCCCCGCCCCCGCCAGTTGACTCTAA 5886

Qy   1544 gGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGl 1564
           ::::|||                             :::   |||     ||||||
Db   5887 AAAGCCT--------------------TTAGAAGAAAAAACAGCACCTCCAGTGACAAA 5925

Qy   1564 nProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCy 1584
                    ::::::|||         ||||||   |||    |||          |||
Db   5926 C---AACTCTGAGATACAAGCCTCGGAGGTGCTGGTAGCT---GCTGACAAGGAAAAGGT 5979

Qy   1584 sLeuProArgLeuGlnLeuLys--------------------AlaProValSer 1595
                |||    :::         |||                    ||||||||
Db   5980 GGCTCCAGTCATTGCTCCCAAAATTACCTCTGTTATTAGCCGGATGCCTGTCAGC 6034
```

RESULT 9

US-10-152-319A-1780

; Sequence 1780, Application US/10152319A
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Higgs, Brandon
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5089-US
; CURRENT APPLICATION NUMBER: US/10/152,319A
; CURRENT FILING DATE: 2002-05-22
; PRIOR APPLICATION NUMBER: US 60/292,335
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/297,523
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,925
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,810
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,807
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/303,808
; PRIOR FILING DATE: 2001-07-10
; PRIOR APPLICATION NUMBER: US 60/315,047
; PRIOR FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/324,928
; PRIOR FILING DATE: 2001-09-27
; PRIOR APPLICATION NUMBER: US 60/330,867
; PRIOR FILING DATE: 2001-11-01
; PRIOR APPLICATION NUMBER: US 60/330,462
; PRIOR FILING DATE: 2001-10-22
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2221
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
;   LENGTH: 15231
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. NM_022401
US-10-152-319A-1780

```
Alignment Scores:
Pred. No.:               1.06e-09      Length:         15231
Score:                   427.00        Matches:        366
Percent Similarity:      33.95%        Conservative:   258
Best Local Similarity:   19.91%        Mismatches:     745
Query Match:             5.29%         Indels:         471
DB:                      6             Gaps:           70
```

US-09-902-432-4 (1-1596) x US-10-152-319A-1780 (1-15231)

```
Qy      3 AlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThrProSer 22
          ||||||    :::|||      |||                       |||:::     |||
Db   4467 GCTGGCTGAGCAACAGCGGGCAGA------------------GGAGCGGGAGCGCCTGGC 4508

Qy     23 GluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspPro 42
                         |||   |||   |||   ||||||:::     |||         |||
Db   4509 CGA-----------GGTGGAGGCCGCGCTGGAGAAGCAGCGGCAGCTGGCTGAGGCCCA 4556

Qy     43 -AlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAs 62
             |||   |||            :::|||         :::      |||
Db   4557 TGCCCAGGCCAAGGCACAGGCCGAGCTGGAGGCACGAGAACTGCAGCGGCGCATGCAGGA 4616

Qy     62 nGlyValAla------GluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGl 80
             |||          |||:::         |||      |||::::::     :::       ||
Db   4617 GGAGGTGACGCGGCGCGAGGAGGCGGCGGTGGACGCACAGCAACAGAAGCGCAGCATCCA 4676
```

```
Qy     80 nGluGlu------------------------------------------GluValValAs 86
Db   4677 AGAGGAGCTGCAGCATCTGCGGCAAAGCTCAGAGGCAGAGATCCAGGCCAAGGCCCAGCA 4736

Qy     86 pGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArgValGluGluMe 106
Db   4737 GGTGGAGGCTGCAGAGCGCAGCCGCATGCGCATTGAGGAAGAGATCCGCGTAGTCCGTCT 4796

Qy    106 tAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGluGluThrSerGl 126
Db   4797 GCAGCTAGAGACAACT---------GAGCGTCAGCGTGGAGGGGCGGAGGATGAGCTGCA 4847

Qy    126 uIleIle------------GluGlnIleProAlaSerGluAsnAsnValGluGluMetVa 142
Db   4848 GGCTCTGCGTGCACGGGCTGAGGAGGCAGAAGCACAGAAGCGGCAGGCTCAGGAGGAAGC 4907

Qy    142 lGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGl 162
Db   4908 CGAGCGCTTGCGGAGGCAGGTGCAGGATGAGAGCCAACGCAAACGGCAGGCGGAGGCCGA 4967

Qy    162 yPheLysPheThrValLysLysAsp---------LysAsnGluLysSerAspThrValGl 179
Db   4968 GCTGGCCCTGCGTGTGAAGGCAGAAGCGGAGGCAGCGCGAGAGAAGCAGCGGGCCCTGCA 5027

Qy    179 nLeuLeuThrValLysLysAspGluGlyGluGlyAlaGlu---------------AlaSe 194
Db   5028 GGCTCTGGATGAACTGAAACTGCAGGCCGAGGAGGCCGAACGGTGGCTGTGCCAAGCCGA 5087

Qy    194 rValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGly----------- 210
Db   5088 GGCAGAGAGGGCTCGCCAAGTGCAGGTAGCCCTGGAGACAGCGCAGCGTAGTGCAGAAGT 5147

Qy    211 -GluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLe 230
Db   5148 GGAGCTGCAGAGCAAGCGTCCGTCCTTTGCAGAGAAGACCGCACAGTTGGAGCGCACGCT 5207

Qy    230 uLysGlnGluGln---SerSerThrGluIleProLeuGlnAlaGlu---------SerAs 246
Db   5208 GCAGGAAGAGCACGTGACAGTGACACAGCTGCGGGAGGAGGCGGAACGGCGGGCACAGCA 5267

Qy    246 pGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGluProTh 266
Db   5268 GCAGGCTGAAGCCGAGCGAGCCCGTGAGGAAGCCGAGCGGGAGCTGGAGCGCTGGCAGCT 5327

Qy    266 rLysSerProGluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLy 286
Db   5328 GAAGGCCAATGAGGCGCTGCGGCTGCGGCTGCAGGCAGAGGAGGTGGCACAGCAGAAGAG 5387

Qy    286 sPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGl 306
Db   5388 CCTG---------------------------GCCCAGGCCGATGCGGAGAAGCAGAAGGA 5420

Qy    306 uAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGl 326
Db   5421 AGAGGCAGAACGGGAAGCCCGGCGGCGGGGCAAGGCAGAGGAGCAGGCCGTGCGGCAGCG 5480

Qy    326 uGluLysGluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGl 346
Db   5481 AGAGCTGGCTGAGCAGGAG---CTGGAGAAGCAGCGGCAGCTGACAGAGGGCACCGCCCA 5537

Qy    346 nAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAs 366
Db   5538 GCAGCGCCTGGCTGCCGAGCAGGAGCTGATTCGCCTGCGGGCAGAGACGGAGCAAGGTGA 5597

Qy    366 pLeuGluAlaSer---SerGluGluLysCysAlaProLeu---------AlaThrGluVa 382
Db   5598 GCATCAGCGGCAGCTGCTGGAGGAAGAGCTGGCCCGGCTACAGCACGAAGCGACAGCAGC 5657

Qy    382 lPheAspGluLysMetGluAlaHisGlnGlu---------ValValAlaGluValHisVa 399
Db   5658 CACACAGAAGCGCCAGGAGCTGGAGGCTGAGCTGGCGAAGGTTCGGGCAGAGATGGAGGT 5717

Qy    399 lSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyVa 419
Db   5718 ACTGCTGGCCAGCAAGGCACGAGCCGAAGAGGAGTCTCGCTCCACCAGTGAA-------- 5769

Qy    419 lValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluVa 439
Db   5770 -------------------------------------------AAGTCCAAGCAGAGGCT 5786

Qy    439 lProGlnGluAla------------GluProAlaGluGluLeuMetLysSerArgGluMe 455
Db   5787 GGAAGCTGAGGCAGGGCGGTTTCGAGAGCTGGCTGAGGAGGCTGCCCGCCTGCGTGCTCT 5846

Qy    455 tCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLe 475
Db   5847 GGCCGAGGAGGCAAGCGGCACCGGGAGTTGGCC-------------GAGGAGGACGCGGC 5894

Qy    475 uProLysHisProGlu------GlyIleValSerGlu---ValGluMetLeuSerSerGl 492
Db   5895 ACGCCAGCGGGCCGAGGCGGACGGAGTGCTTACGGAGAAGCTGGCTGCCATCAGTGAGGC 5954

Qy    492 nGluArgIleLys---------------------------------ValGlnGlySerPr 501
Db   5955 CACAAGGCTCAAGACGGAGGCAGAGATTGCACTCAAAGAGAAGGAGGCCGAGAACGAGCG 6014

Qy    501 oLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGl 521
Db   6015 CCTGAGGCGCCTGGCTGAAGATGAGGCCTTCCAGCGGCGCCGGCTGGAGGAGCAGGCAGC 6074

Qy    521 yLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSe 541
Db   6075 ACAGCACAAGGCAGACATAGAGGAGCGCCTGGCCCAG---------CTGCGCAAGGCATC 6125

Qy    541 rProGluSerAlaAspGluGlnLysGly------GluSerSerAlaSerSerProGluGl 559
Db   6126 CGAGAGCGAGCTGGAGCGACAGAAGGGGTTGGTGGAGGATACCCTGCGGCAGCGGCGGCA 6185

Qy    559 uProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAl 579
Db   6186 GGTGGAGGAGGAGATCATGGCTCTGAAGGCGAGCTTCGAGAAGGCCGCGGCTGGCAAGGC 6245

Qy    579 aGlu------------------------GluGlyThrThrSerAspGlyGluLysLysAr 591
Db   6246 AGAACTGGAGCTGGAGCTTGGCCGCATCCGCAGCAATGCCGAGGACACCATGCGCAGCAA 6305

Qy    591 gGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValAr 611
Db   6306 GGAGCTGGCCGAGCAGGAGGCAGCGCGGCAGCGGCAGTTGGCAGCTGAGGAGGAGCAGAG 6365

Qy    611 gArgProSerGluSerAspLys---------------------GluGluGluLeuGluLy 624
Db   6366 GCGCCGGGAAGCCGAGGAGCGGGTGCAGAGGAGCCTGGCAGCGGAGGAGGAAGCCGCACG 6425

Qy    624 sValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluVa 644
Db   6426 GCAGCGCAAGGTCGCACTGGAGGAAGTCGAGCGGCTCAAGGCCAAGGTT---GAGGAAGC 6482

Qy    644 lLysThrValGlyGluGluGlnLysProGluGluProLys-------------------- 657
Db   6483 GCGGCGCCTGCGAGAGCGAGCTGAGCAGGAGTCTGCGAGGCAGCTGCAGCTGGCCCAGGA 6542

Qy    658 ----------ArgArgValAspThrSerValSerTrpGluAlaLeuIleCys-------- 671
Db   6543 GGCTGCCCAGAAACGGCTGCAGGCGGAGGAGAAGGCGCACGCCTTTGTGGTGCAGCAGCG 6602

Qy    672 ----------------------------------------------------ValGlySe 674
Db   6603 AGAAGAGGAGCTGCAGCAGACTCTTCAGCAAGAGCAGAACATGCTGGAGCGGCTGCGGAG 6662

Qy    674 rSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGly---------GlyPr 691
Db   6663 CGAGGCAGAGGCAGCGCGGCGAGCTGCTGAGGAGGCGGAGGAGGCCCGGGAGCAGGCAGA 6722

Qy    691 oArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGl 711
```

```
                 |||                         :::   |||||||||   :::   |||::::::
Db      6723 ACGTGAGGCAGCGCAGTCTAGGAAGCAAGTGGAAGAGGCCGAGCGGCTGAAGCAGTCG-- 6780

Qy       711 yThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPr 731
                              |||      |||    |||   ||||||||||      ::::::
Db      6781 ----------------GCAGAGGAGCAGGCTCAGGCCCAGGCCCAGGCGCAGGCGGCTGC 6824

Qy       731 oGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysAr 751
              |||
Db      6825 AGAG-------------------------------------------------------- 6828

Qy       751 gLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSe 771
                              |||      |||    :::|||   |||         :::
Db      6829 ----------------AAACTGCGCAAGGAAGCGGAGCAGGAGGCGGCGCGTCGGGCCCA 6872

Qy       771 rValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLy 791
                ||||||   :::   :::   :::   :::   :::   :::                ||
Db      6873 GGCGGAGCAGGCTGCGTTGAAACAGAAGCAGGCAGCCGACGCGGAGATGGAGAAGCACAA 6932

Qy       791 sLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrVa 811
             |||||||                   :::::::::|||      :::||||||      |||::
Db      6933 GAAGTTTGCAGAGCAGACGCTACGGCAGAAGGCTCAGGTAGAGCAGGAGCTGACCACGCT 6992

Qy       811 lGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValValPr 831
             :               :::::::::   |||   |||   :::      :::::::::
Db      6993 GAGG------------CTGCAGCTCGAGGAGACCGAC---CACCAGAAGAGCATCCTGGA 7037

Qy       831 oLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---AlaGlnGlyAsnThrGl 850
                      |||          ::::::   |||      |||   |||:::       :::::
Db      7038 T---GAGGAGCTGCAGCGGCTAAAGGCTGAGGTAACAGAGGCAGCCCGGCAGCGTAGCCA 7094

Qy       850 uLeuProGlnLeuLeuGlyAlaValTyrValSer---GluGluLeuSerLysThrLeuVa 869
             ::::   :::   :::|||   :::|||   |||||||||   |||
Db      7095 GGTAGAGGAGGAGCTCTTCTCTGTCCGCGTGCAGATGGAGGAGCTGGGCAAA-------- 7146

Qy       869 lHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluAr 889
                                             :::|||            |||   ||
Db      7147 ------------------------------CTCAAGGCTCGCATTGAAGCTGAAAACCG 7175

Qy       889 gSerProSerTrpIleSerAlaSerValThrGluPro---LeuGluHisThrAlaGlyGl 908
             |:::                           |||:::      ||||||      |||    ::
Db      7176 GGCACTCATCCTTCGTGACAAGGACAACACACAGCGCTTCCTGGAGGAGGAGGCCGAGAA 7235

Qy       908 uAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVa 928
             :                  ||||||          :::      |||:::|||
Db      7236 GATGAAACAGGTGGCAGAGGAAGCTGCACGGTTGAGCGTAGCTGCCCAGGAGGCAGCAAG 7295

Qy       928 lLeuThrGln------------ThrLeuProGluGlyLysAspAlaHisAspAspMetVa 944
               |||   |||            |||      :::   :::         :::      |||::
Db      7296 GCTGCGGCAGCTAGCCGAGGAGGACCTGGCCCAGCAGCGGGCCCTGGCGGAGAAGATGCT 7355

Qy       944 lThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLe 964
             :            ::::::            ||||||            :::|||     ||
Db      7356 GAAGGAGAAGATGCAGGCGGTGCAGGAAGCCACAAGGCTCAAGGCTGAGGCTGAGCTGCT 7415

Qy       964 uArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAl 984
             |:::      ::::::                  |||:::|||            :::      ||
Db      7416 GCAGCAGCAGAAGGAGCTG------------GCACAGGAGCAGGCCCGGCGGCTGCAGGC 7463

Qy       984 aValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVa 1004
             |                            |||      :::   |||      :::   |||
Db      7464 G------------------------GACAAGGAGCAAATGGCTCAGCAGTTGGTAGAGGA 7499

Qy      1004 lGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAl 1024
                         |||   |||   |||   |||||||||   |||         :::      ||
Db      7500 GACACAGGGTTTCCAGCGGACCCTGGAGGCTGAGCGGCAGCGGCAGCTAGAAATGAGCGC 7559

Qy      1024 aValAlaAspLysValLys----------------------------------------- 1030
             |    |||:::::::::|||
Db      7560 AGAGGCTGAACGCCTCAAGTTGCGCATGGCTGAGATGAGCCGGGCTCAGGCCCGTGCAGA 7619

Qy      1031 -GluGluSerGln----------------------------------------------- 1034
               |||:::::::|||
Db      7620 GGAGGATGCCCAGCGCTTCCGGAAGCAGGCTGAAGAGATCGGCGAAAAGCTGCACCGCAC 7679

Qy      1035 ----------------------ValProAlaThrGlnThr------ValGlnArgThrGl 1045
                                   |||         ||||||        :::||||||
Db      7680 TGAACTCGCTACACAGGAGAAGGTGACATTGGTGCAGACTCTCGAGATCCAGCGACAGCA 7739

Qy      1045 ySer---------------LysAlaLeuGluLysValGluGluValGluGluAspSerGl 1060
              |||                :::      |||   :::   |||:::|||   :::   ||
Db      7740 GAGTGACCAAGATGCCGAGCGTCTGAGGGAGGCCATTGCTGAGCTGGAGCGTGAGAAGGA 7799

Qy      1060 u--------------------------ValLeuAlaSerGluLysGluLysAspValMe 1071
             |                              |||   ||||||:::   :::   |||
Db      7800 GAAGCTCAAGCAGGAGGCGAAGTTACTGCAGCTCAAGTCTGAGGAGATGCAGACTGTGCA 7859

Qy      1071 tProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGl 1091
                :::         :::||||||      |||         :::         ||||||
Db      7860 GCAGGAGCAGATACTGCAGGAGACACAGGCCCTGCAGAAGAGCTTTCTCTCTGAGAAGGA 7919

Qy      1091 yGlnAlaThrProGluSerLeuGluValProGluValThrAlaAspValAspHisValAl 1111
                                 :::   :::         |||   ::::::         :::
Db      7920 CAGCTTGCTGCAACGCGAACGCTTCATCGAGCAGGAGAAGGCCAAGCTGGAGCAGCTTTT 7979

Qy      1111 aThrCysGlnValIleLysLeuGlnGlnLeuMetGlu------------------GlnAl 1125
                   :::|||   |||   :::||||||   |||                        |||
Db      7980 CCAGGACGAGGTGGCAAAAGCAAAACAGCTGCAGGAGGAGCAGCAGCGGCAGCAGCAGCA 8039

Qy      1125 aValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLe 1145
              :::      |||      |||   :::            |||
Db      8040 GATGGAGCAGGAAAAGCAGGAGCTGGTGGCCAGCATGGAG---------GAGGCCCGGAG 8090

Qy      1145 uAlaAspSerAspThrAlaAspGlyThr------------------------------ 1154
                         :::      :::|||
Db      8091 GCGGCAGCGTGAGGCAGAGGAGGGTGTGAGGCGCAAGCAAGAGGAACTGCAGCGTCTGGA 8150

Qy      1155 ---------GlnGlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaVa 1171
                      ||||||   |||   :::   ::::::::::::::
Db      8151 GCAGCAGCGGCAGCAGCAGGAGAAACTACTGGCAGAGGAGAACCAGAGGCTGCGGGAGCG 8210

Qy      1171 lArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerTh 1191
                |||         |||   ||||||   |||      ||||||   :::||
Db      8211 GCTGCAGCGCCTGGAGGAAGAGCACCGAGCTGCGTTGGCGCACTCTGAGGAGATCGCCAC 8270

Qy      1191 r-----------------------LeuProAsnAsnValProAlaGlnGluGluHisGl 1203
             |                         |||||||||
Db      8271 CTCCCAGGCTGCTGCCACAAAAGCACTGCCCAAT------------------------- 8304

Qy      1203 yGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaVa 1223
                      |||||||||   |||:::               |||
Db      8305 ----------GGCCGCGACGCACTTGATGGCCCCTCCATGGAG---------------- 8337

Qy      1223 lProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGl 1243
                      |||:::   |||         ||||||         :::      ::
Db      8338 ----------GCCGAGCCCGAGTACACCTTTGAGGGATTACGTCAGAAGGTGCCAGCTCA 8387

Qy      1243 uLysValLysGlu------------GluGlnGluValPheValHisSerGlyProAsnSe 1259
             :::::::::|||               |||
Db      8388 GCAGCTACAGGAAGCAGGCATTCTGAGCATGGAGGAACTGCAGCGTTTGACACAGGTCA 8447

Qy      1259 rGlnLysAlaAlaAspValThrTyrAspSerGluVal--------------------Me 1272
                      |||::::::|||         :::|||
Db      8448 CACCACGGTGGCTGAGCTCACGCAGCGGGAAGATGTGCGCCACTACCTGAAGGGCGGCAG 8507

Qy      1272 tGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSer------------- 1287
                  :::||||||         |||   :::   |||      |||
Db      8508 CAGCATCGCAGGATTGCTCCTGAAGCCCACCAATGAGAAACTGAGTGTCTACACAGCCCT 8567

Qy      1288 ------------------------------LeuSerLeuGluGluGlyGluMetGluTh 1297
                                           |||   |||   |||   :::      ::
Db      8568 ACAGCGGCAGCTGCTCAGCCCTGGAACAGCCCTTATCTTACTTGAGGCCCAGGCAGCCTC 8627

Qy      1297 r------------AspValGluLysGluLysArgGluThrLysProGluGlnValSerGl 1313
             :               |||      :::   :::|||   |||         |||   |||   ||
Db      8628 GGGCTTCCTGCTGGACCCTGTCCGGAACCGGCGGCTGACGGTCAATGAGGCTGTGAAGGA 8687
```

```
Qy   1313 u------------------------------------------------------------ 1313
Db   8688 GGGTGTGGTGGGTCCCGAGCTGCACCACAAGCTGCTGTCAGCTGAGCGTGCCGTCACTGG 8747

Qy   1314 ----------------GluGlyGluGlnGluThrAlaAlaProGluHisGlu-------- 1325
Db   8748 CTACAAGGACCCTTACACAGGAGAACAGAT-CTCTCTCTTCCAGGCCATGAAGAAGGACC 8806

Qy   1326 ----------GlyThrTyrGlyLysProValLeuThrLeuAsp----------------- 1336
Db   8807 TCATTGTCAGGGACCATGGCATCCGCCTGCTGGAAGCCCAGATCGCCACAGGTGGCATCA 8866

Qy   1337 ------------MetProSerSerGluArgGlyLysAlaLeuGlySerLeu-------- 1349
Db   8867 TTGACCCTGTACACAGCCACCGTGTTCCCGTGGACGTGGCCTACCAGCGTGGCTACTTCG 8926

Qy   1350 ----GlyGlySerProSer-Leu------ProAspGlnAspLysAlaGlyCysIleGluV 1366
Db   8927 ATGAGGAGATGAACCGTGTGCTGGCTGACCCAAGCGATGACACCAAGGGCTTCTTTGACC 8986

Qy   1366 alGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValI 1386
Db   8987 CCAACACTCAC--------------------------------------GAGAACCTCA 9007

Qy   1386 leGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuL 1406
Db   9008 CGTACCTGCAGCTGCTGGAGCGCTGTGTGGAGGACCCCGAG---ACAGGCCTGCGCCTCC 9064

Qy   1406 euProAlaGluLysSerSerAlaThrGlyGlyHis-Trp---ThrLeuGlnHisAlaGlu 1424
Db   9065 TGCCACTCACAGACAAGGCTGCCAAGGGTGGTGAGCTGGTGTACACTGACACGGAGGCCC 9124

Qy   1425 AspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrPro 1444
Db   9125 GTGACGTCTTCGAAAAGGCCACAGTGTCTGCACCATTCGGCAAGTTCCAGGGCAAGACCG 9184

Qy   1445 AlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGlu 1464
Db   9185 TGACCATCTGGGAGA-------------------------TCATCAACTCAGAGTACT 9217

Qy   1465 ArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThr 1484
Db   9218 TCACAGCGGAGCAGCGACGGGACCTGCTCCGGCAGTTCCGCACG-GGCCGC------ATC 9270

Qy   1485 AlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsn 1504
Db   9271 ACGGTGGAGAAGATCATCAAGATTGTCATCACGGTGGTAGAGGAACACGAGCGGAAGGGC 9330

Qy   1505 LysIle-ValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAl 1524
Db   9331 CAGCTCTGCTTTG------------------------------------------------ 9343

Qy   1524 aProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerAr 1544
Db   9344 ------------------------AGGGCCTCCGTGCCCTTGTGCCTGCTGCAGAGC-- 9376

Qy   1544 gGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLys 1559
Db   9377 ------------TGCTGGACA---GTGGAGTCATCAGTCATGAAG 9406
```

RESULT 10
US-09-724-676-25032
; Sequence 25032, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25032
;   LENGTH: 14335
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-25032

Alignment Scores:

Pred. No.:	2.38e-09	Length:	14335
Score:	418.00	Matches:	334
Percent Similarity:	35.65%	Conservative:	255
Best Local Similarity:	20.22%	Mismatches:	701
Query Match:	5.18%	Indels:	364
DB:	5	Gaps:	61

US-09-902-432-4 (1-1596) x US-09-724-676-25032 (1-14335)

```
Qy      5 SerSerThrGluGlnArgSerProGlu-GlnProAlaGlySerAspThrProSerGluLe 24
Db   4187 AGCAGCAGCGGGCAGAGGAGCGCGAGCGGCTGGCCGAGGTGGAGGCCGCGCTGGAGAAGC 4246

Qy     24 uValLeuSerGlyHisGly------ProAlaAlaGluAlaSerGlyAlaAlaGlyAspPr 42
Db   4247 AGCGGCAGCTGGCCGAGGCGCACGCCCAGGCAAAGGCACAGGCGGAGCGGGAGGCGAAGG 4306

Qy     42 oAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAs 62
Db   4307 AGCTGCAGCAGCGCATGCA------------GGAGGAGGTGGTGCGGCGGGAGGAGGCGG 4354

Qy     62 nGly-ValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGlnGluG 82
Db   4355 CGGTGGACGCGCAGCAGCAGAAGCGCAGCATTCAGGAGGAGCTGCAGCAGCTGCGGCAGA 4414

Qy     82 luGluValValAspGluAspValGlyGlnArgGluSerGlu------------------- 95
Db   4415 GCTCGGAGGCGGAGATCCAGGCCAAGGCCCGGCAGGCAGAGGCGGCTGAGCGCAGCCGGC 4474

Qy     96 --AspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluA 115
Db   4475 TGCGCATCGAGGAGGAGATCCGCGTGGTGCGCCTGCAGTTGGAGGCCACCGAGCGCCAGC 4534

Qy    115 spIleThrLysAspGlyGlnGluGluThrSerGluIle---IleGluGlnIleProAlaS 134
Db   4535 GTGGCGGGGCTGAGGGGGAGCTGCAGGCACTGCGTGCACGGGCGGAGGAGGCTGAGGCAC 4594

Qy    134 erGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyP 154
Db   4595 AAAAGCGACAGGCGCAGGAGGAGGCCGAGCGCTTGCGGAGGCAGGTGCAGGACGAGAGCC 4654

Qy    154 heLysLysValPheLysPheValGlyPheLysPheThrValLysLysAsp---------L 171
Db   4655 AGCGTAAGCGGCAGGCGGAGGTGGAGCTGGCCTCGCGCGTGAAGGCCGAGACCGAGGCGG 4714

Qy    171 ysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyA 191
Db   4715 CGCGCGAGAAGCAGCGGGCCCTGCAGGCCCTGGAGGAGCTGCGGCTGCAGGCGGAGGAGG 4774

Qy    191 laGlu---------------AlaSerValGlyAlaGlyAspHisGlnGluProSerValG 206
Db   4775 CGGAGCGGCGCCTGCGGCAGGCCGAGGTGGAGCGAGCGCGGCAGGTACAGGTGGCCCTGG 4834

Qy    206 luThr------------AlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnS 222
Db   4835 AGACGGCGCAGCGCAGTGCAGAGGCGGAGCTGCAGAGCAAACGCGCCTCCTTCGCCGAGA 4894

Qy    222 erThrGluLysGlnGluGlyThrLeuLysGlnGluGln---SerSerThrGluIleProL 241
Db   4895 AGACGGCACAGCTGGAGCGCTCCCTGCAGGAGGAACACGTGGCTGTGGCACAGCTGCGGG 4954

Qy    241 euGlnAlaGlu---------SerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyG 258
Db   4955 AGGAGGCTGAGCGGCGGGCACAGCAGCAGGCCGAGGCCGAGCGGGCGCGCGAGGAGGCAG 5014

Qy    258 luGluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnS 278
```

```
Db   5015 AGCGGGAGCTGGAGCGCTGGCAGCTCAAGGCCAACGAGGCGCTACGGCTGCGGCTGCAGG 5074
Qy    278 erGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysL 298
Db   5075 CGGAGGAGGTGGCGCAGCAGAAGAGCCTG--------------------------GCGC 5107
Qy    298 ysThrSerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluG 318
Db   5108 AGGCCGAGGCTGAGAAGCAGAAGGAGGAGGCGGAGCGCGAGGCGCGGCGGCGCGGCAAGG 5167
Qy    318 lnGluAlaGluLysValAspGluGluLysGluLysThrGluProAlaSerGluGluG 338
Db   5168 CGGAGGAGCAGGCCGTCCGGCAGCGGGAGCTGGCTGAACAAGAG---CTGGAGAAGCAGC 5224
Qy    338 lnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluL 358
Db   5225 GGCAGCTGGCGGAAGGCACCGCGCAGCAGCGCCTGGCCGCGGAGCAGGAGTTGATCCGGC 5284
Qy    358 euProLeuGluAspGlnValGlyAspLeuGluAlaSer---SerGluGluLysCysAlaP 377
Db   5285 TGCGGGCCGAGACGGAGCAGGGGGAGCAGCAGCGGCAGCTGCTGGAGGAGGAGCTGGCCC 5344
Qy    377 roLeu-----------------AlaThrGluValPheAspGlu--------------- 385
Db   5345 GGCTGCAGCGTGAGGCGGCTGCAGCCACGCAGAAACGGCAGGAGCTGGAAGCCGAGCTGG 5404
Qy    386 --LysMetGluAlaHisGlnGluValVal--------------------AlaGluValH 398
Db   5405 CCAAGGTGCGGGCCGAGATGGAGGTGCTGCTGGCCAGCAAGGCGAGGGCTGAGGAGGAGT 5464
Qy    398 isValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyG 418
Db   5465 CGCGCTCCACCAGCGAGAAGTCCAAGCAGAGGCTG-----------GAGGCCGAGGCCG 5512
Qy    418 lyValValValGluGlyThrGlyGlu---------------SerLeuProProGlu---- 431
Db   5513 GCCGGTTCCGCGAGCTGGCCGAGGAGGCCGCCCGCCTGCGTGCCCTGGCGGAAGAGGCCA 5572
Qy    432 -----------LysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluG 448
Db   5573 AGCGGCAGCGGCAGCTGGCCGAGGAAGACGCGGCGCGGCAGCGGGCCGAG---GCGGAGC 5629
Qy    448 luLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspL 468
Db   5630 GGGTGCTTGCGGAGAAGCTGGCCGCCATCGGC---------------GAGGCCACGCGGC 5674
Qy    468 euSerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluM 488
Db   5675 TCAAGACGGAGGCGGAGATCGCGCTCAAGGAGAAGGAGGCGGAGAACGAGCGCCTGCGGC 5734
Qy    488 etLeuSerSerGlnGlu---------------ArgIleLysValGlnGlySerProLeuL 503
Db   5735 GGCTGGCGGAGGACGAGGCCTTCCAGCGGCGGCGGCTGGAGGAGCAGGCCGCGCAACACA 5794
Qy    503 ysLysLeuPheSerSer------SerGlyLeuLysLysLeuSerGly---------LysL 518
Db   5795 AGGCTGACATCGAGGAGCGCCTGGCCCAGCTGCGCAAGGCATCGGACAGCGAGCTGGAGC 5854
Qy    518 ysGlnLysGlyLysArgGlyGly--------------------GlyGlyAspGluGluP 531
Db   5855 GGCAGAAGGGGCT-GGTGGAGGACACGCTGAGGCAGCGGCGGCAGGTGGAGGAAGAGATC 5913
Qy    531 roGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln------L 549
Db   5914 CTGGC-----------GCTGAAGGCGAGCTTCGAGAAGGCGGCCGCTGGCAAGGCGGAG 5961
Qy    549 ysGlyGluSerSerAlaSerSerProGluGlu-ProGluGluThr---ThrCysLeuGlu 567
Db   5962 CTGGAGCTGGAGCTGGGACGCATCCGCAGCAACGCGGAGGACACGCTGCGCAGCAAGGAG 6021
Qy    568 LysGlyProLeuGluAlaProGlnAsp-------------------------------- 576
Db   6022 CAGGCCGAGCTGGAGGCCGCGAGGCAGCGGCAGCTGGCGGCGGAGGAGGAGCGGCGGCGC 6081
```

```
Qy    577 GlyGluAlaGluGluGlyThrThrSerAsp--------------------GlyGluLys 589
Db   6082 CGTGAGGCTGAGGAGCGCGTGCAGAAGAGCCTGGCGGCCGAGGAGGAGGCCGCACGGCAG 6141
Qy    590 LysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArg 609
Db   6142 CGGAAGGCGGCGCTGGAGGAAGTCGAGCGGCTGAAAGCCAACGTGGAGGAGGCGCGGCGC 6201
Qy    610 ValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThr 629
Db   6202 CTGCGGGAGCGAGCGGAGCAGGAGTCGGCGCGGCAGCTGCAGCTGGCCCAGGAGGCCGCC 6261
Qy    630 LeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGlu 649
Db   6262 CAGAAGCGGCTG--------------CAGGCGGAAGAGAAGGCACACGCCTTCGCGGTG 6306
Qy    650 GluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeu 669
Db   6307 CAGCAGAAGGAGCAGGAGCTACAGCAGACGCTGCAGCAGGAGCAGAGC------GTGCTG 6360
Qy    670 IleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGly 689
Db   6361 GACCAGCTGCGCGGCGAGGCGGAGGCGGCCCGGCGGGCGGCTGAGGAGGCGGAGGAGGCC 6420
Qy    690 ---------GlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLys 706
Db   6421 CGGGTGCAGGCGGAGCGTGAGGCGGCGCAGGCCCGGCGGCAGGTGGAAGAGGCCGAGCGG 6480
Qy    707 AspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGln 726
Db   6481 CTGAAGCAGTCG-----------------GCAGAGGAGCAGGCACAGGCCCGGGCTCAG 6522
Qy    727 GlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrp 746
Db   6523 GCACAGGCGGCTGCAGAG------------------------------------------ 6540
Qy    747 GluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLys 766
Db   6541 ------------------------------AAGCTGCGCAAGGAGGCCGAGCAAGAGGCG 6570
Qy    767 AlaGluAspSerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSer 786
Db   6571 GCGCGGCGGGCACAGGCGGAGCAGGCGGCCCTGCGGCAGAAGCAGGCAGCTGACGCGGAG 6630
Qy    787 TrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGln 806
Db   6631 ATGGAGAAGCATAAGAAATTCGCCGAGCAGACGCTGCGGCAGAAGGCGCAGGTGGAGCAG 6690
Qy    807 GluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnVal 826
Db   6691 GAGCTGACAACACTGCGG------------CTGCAGCTGGAGGAGACCGAC--------- 6729
Qy    827 ProAlaValValProLeuSerGluTyrAsnAlaValGluArgGlu------------Lys 842
Db   6730 -----------------CACCAGAAGAACCTGCTGGACGAGGAGCTGCAGCGGCTGAAG 6771
Qy    843 MetGlu-----------AlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaVal 858
Db   6772 GCGGAGGCCACGGAGGCCGCACGCCAGCGCAGCCAGGTGGAGGAGGAGCTCTTCTCGGTG 6831
Qy    859 TyrValSer---GluGluLeuSerLysThrLeuValHisThrValSerValAlaValIle 877
Db   6832 CGCGTGCAGATGGAGGAGCTGAGCAAG-------------------------------- 6858
Qy    878 AspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSer 897
Db   6859 ------CTCAAGGCACGCATCGAGGCTGAGAACCGCGCACTCATCTTGCGTGACAAGGAC 6912
Qy    898 ValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProVal---GluGluVal 916
Db   6913 AATACGCAGCGCTTCCTGCAGGAGGAGGCTGAGAAGATGAAGCAGGTGGCGGAGGAGGCC 6972
```

```
Qy     917 ThrGluLysAspIleIleAlaGluGluThrProValLeuThrGln-----------Thr 932
                       :::   |||:::|||               |||    |||
Db    6973 GCGCGGCTGAGTGTGGCGGCCCAAGAGGCTGCGCGACTGCGGCAGCTGGCAGAGGAGGAC 7032

Qy     933 LeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAspPheThrSer 952
           |||   :::   :::          :::   |||:::       ::::::
Db    7033 CTGGCACAGCAGCGGGCCTTGGCAGAGAAGATGCTCAAGGAGAAGATGCAGGCGGTGCAG 7092

Qy     953 GluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAla 972
           ||||||                      :::|||   |||:::   ::::::
Db    7093 GAGGCCACGCGACTCAAGGCTGAGGCGGAACTGCTGCAGCAGCAGAAGGAGCTTGCGCAG 7152

Qy     973 SerGlyAlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerPro 992
                 |||               :::
Db    7153 GAGCAGGCGCGGCGGCTGCAGGAG------------------------------------ 7176

Qy     993 AspThrThrGluGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGlu 1012
           |||      :::   |||      :::   |||         |||         |||
Db    7177 GACAAGGAGCAGATGGCGCAGCAGCTGGCGGAGGAGACGCAGGGCTTCCAGCGGACGCTG 7236

Qy    1013 GluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLys------ 1030
           |||   |||||||||   |||      :::   |||   |||:::::::::|||
Db    7237 GAGGCCGAGCGGCAGCGGCAGCTGGAGATGAGCGCTGAGGCTGAGCGCCTCAAGCTGCGT 7296

Qy    1031 ------------GluGluSerGlnValProAlaThrGlnThrValGlnArgThrGlySer 1046
                          :::|||      |||   :::         ||||||
Db    7297 GTGGCCGAGATGAGCCGAGCCCAGGCCCGCGCTGAGGAGGACGCCCAGCGCTTCCGGAAG 7356

Qy    1047 LysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys 1066
           :::|||   |||         |||::::::                  ||||||:::::::::
Db    7357 CAGGCGGAGGAGATCGGTGAGAAGCTGCACCGCACGGAG------CTCGCCACCCAGGAG 7410

Qy    1067 GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGln 1086
           :::         ::::::            :::|||            ::::::|||   |||:::
Db    7411 AAGGTGACCCTGGTGCAGACACTGGAGATCCAGCGACAGCAGAGTGACCATGATGCCGAG 7470

Qy    1087 GlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrAlaAsp 1106
                                                   :::   |||      |||:::
Db    7471 -----------------------------------CGCCTGCGGGAGGCCATCGCTGAG 7494

Qy    1107 ValAspHis------ValAlaThrCysGlnValIleLysLeuGlnGlnLeu---MetGlu 1123
           ::::::                           |||      ||||||   ||||||      |||
Db    7495 CTGGAGCGTGAGAAGGAGAAGCTCCAACAGGAGGCCAAACTGCTGCAGCTCAAGTCTGAG 7554

Qy    1124 GlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThr 1143
           :::            :::      :::   |||   ::::::::            :::
Db    7555 GAGATGCAGACGGTGCAGCAGGAGCAGCTGCTGCAGGAGACGCAGGCCCTGCAGCAAAGC 7614

Qy    1144 ProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGln 1163
              |||::::::   |||:::               |||::::::      |||:::   :::
Db    7615 TTCCTCTCTGAAAAGGACAGCCTG------CTACAGCGGGAGCGCTTCATCGAGCAGGAG 7668

Qy    1164 AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThr 1183
              :::|||                  |||:::   ::::::|||   |||
Db    7669 AAGGCCAAGCTG------------------GAGCAGCTCTTCCAGGACGAGGTGGCCAAG 7710

Qy    1184 AlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGly 1203
           ||||||:::                                    :::||||||
Db    7711 GCACAGCAG--------------------------------CTGCGTGAGGAG------ 7731

Qy    1204 GluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaVal 1223
           ::::::      :::      :::|||      :::|||   |||   |||:::
Db    7732 CAGCAGCGGCAGCAGCAGCAGATGGAGCAGGAACGGCAGCGGCTGGTGGCCAGCATGGAG 7791

Qy    1224 ProValLeuAlaLysThrGluValGlyGlnGluGly------------------GluVal 1237
                       :::         :::||||||                  |||:::
Db    7792 GAGGCGCGGCGGCGGCAGCATGAGGCCGAGGAGGGCGTGCGGCGCAAGCAGGAGGAGCTG 7851

Qy    1238 AspTrpLeuAspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyPro 1257
              |||:::   ::::::      ::::::::::::|||:::
Db    7852 CAGCAGCTGGAGCAGCAGCGG---CGGCAGCAGGAGGAGCTGCTGGCTGAGGAGAACCAG 7908

Qy    1258 AsnSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCys 1277
                 ::::::      :::      :::                  :::|||
Db    7909 AGGCTGCGTGAGCAGCTGCAGCTCCTGGAGGAGCAGCACCGGGCCGCGCTGGCGCACTCA 7968

Qy    1278 GlnGluLysGluSerThrGluValGlnSerLeuSer----------------------- 1289
           :::|||         :::::::::|||   :::
Db    7969 GAGGAGGTCACTGCCTCGCAGGTGGCTGCCACAAAGACCCTGCCCAATGGCCGGGATGCA 8028

Qy    1290 ------------LeuGluGlyGluMetGluThrAspValGluLysGluLysArgGlu 1305
                                  |||   |||   |||         :::      :::|||:::
Db    8029 CTTGATGGCCCCGCGGCAGAGGCAGAGCCGGAGCACAGCTTCGATGGCCTGCGGCGGAAG 8088

Qy    1306 ThrLysProGluGlnValSerGluGluGlyGluGlnGluThr----------------- 1319
                 :::::::::   |||   |||      |||
Db    8089 GTGTCAGCTCAGAGGCTGCAGGAGGCCGGCAT-CCTGAGTGCGGAGGAGCTGCAGCGGTT 8147

Qy    1320 ---AlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetPro 1338
              |||   |||   |||   |||
Db    8148 GGCGCAGGGCCACACCACGGTGGA----------------------------------- 8171

Qy    1339 SerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGln 1358
                    |||   :::      |||            |||:::   |||   |||
Db    8172 ---------CGAGCTCGCACGGCGGGAAGACGTGCGCCACTACCTGCAGGGCCGCAGCAG 8222

Qy    1359 AspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThr 1378
              :::         |||
Db    8223 TATCGCAGGGCTGTTGCTGAAGGCCAC-----------------------CAATGAGAA 8258

Qy    1379 AlaGluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerPro 1398
           ||||||   :::   :::         |||               |||      |||      |||
Db    8259 GCTGAGTGTTTACGCCGCCCTGCAGAGGCAGCTGCTGAGTCCCGGCACGGCCCTCATCCT 8318

Qy    1399 Glu-------------------------------------------------------- 1399

Db    8319 GCTGGAGGCGCAGGCGGCCTCAGGCTTCCTGCTGGACCCTGTGCGGAACCGGCGGCTGAC 8378

Qy    1400 -----------------------CysValGlyAlaHisLeuLeuProAlaGluLysSer 1411
                                  |||   |||            |||
Db    8379 CGTCAACGAGGCTGTGAAGGAGGGTGTGGTGGGCCCCGAGCTGCACCACAAGCTGCTGTC 8438

Qy    1412 SerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyPro 1431
                          ||||||                              |||
Db    8439 GGCCGAGCGCGCCGTCACTGGCTACAAGGACCCCTACACTGGCCAGCAGATCTCTCTCTT 8498

Qy    1432 GluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHis 1451
              |||   ||||||                  |||               |||
Db    8499 CCAAGCCATGCAGAAGGG-----------------CCTCATCGTCCGGGAGCACGGCAT 8540

Qy    1452 ProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLys 1471
           |||
Db    8541 CCG-------------------------------------------------------- 8543

Qy    1472 ProAspAlaGlyProAspAlaAspGlyLys 1481
           |||      |||||||||      |||:::
Db    8544 CCTGCTGGAGGCCCAGATCGCCACGGGCGG 8573
```

RESULT 11
US-09-724-676A-25032
; Sequence 25032, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 25032
; LENGTH: 14335
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-724-676A-25032

Alignment Scores:

```
Pred. No.:              2.38e-09      Length:          14335
Score:                  418.00        Matches:         334
Percent Similarity:     35.65%        Conservative:    255
Best Local Similarity:  20.22%        Mismatches:      701
Query Match:            5.18%         Indels:          364
DB:                     5             Gaps:            61

US-09-902-432-4 (1-1596) x US-09-724-676A-25032 (1-14335)

Qy        5 SerSerThrGluGlnArgSerProGlu-GlnProAlaGlySerAspThrProSerGluLe 24
Db     4187 AGCAGCAGCGGGCAGAGGAGCGCGAGCGGCTGGCCGAGGTGGAGGCCGCGCTGGAGAAGC 4246

Qy       24 uValLeuSerGlyHisGly------ProAlaAlaGluAlaSerGlyAlaAlaGlyAspPr 42
Db     4247 AGCGGCAGCTGGCCGAGGCGCACGCCCAGGCAAAGGCACAGGCGGAGCGGGAGGCGAAGG 4306

Qy       42 oAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAs 62
Db     4307 AGCTGCAGCAGCGCATGCA-----------GGAGGAGGTGGTGCGGCGGGAGGAGGCGG 4354

Qy       62 nGly-ValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGlnGluG 82
Db     4355 CGGTGGACGCGCAGCAGCAGAAGCGCAGCATTCAGGAGGAGCTGCAGCAGCTGCGGCAGA 4414

Qy       82 luGluValValAspGluAspValGlyGlnArgGluSerGlu------------------ 95
Db     4415 GCTCGGAGGCGGAGATCCAGGCCAAGGCCCGGCAGGCAGAGGCGGCTGAGCGCAGCCGGC 4474

Qy       96 --AspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluA 115
Db     4475 TGCGCATCGAGGAGGAGATCCGCGTGGTGCGCCTGCAGTTGGAGGCCACCGAGCGCCAGC 4534

Qy      115 spIleThrLysAspGlyGlnGluGluThrSerGluIle---IleGluGlnIleProAlaS 134
Db     4535 GTGGCGGGGCTGAGGGGGAGCTGCAGGCACTGCGTGCACGGGCGGAGGAGGCTGAGGCAC 4594

Qy      134 erGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyP 154
Db     4595 AAAAGCGACAGGCGCAGGAGGAGGCCGAGCGCTTGCGGAGGCAGGTGCAGGACGAGAGCC 4654

Qy      154 heLysLysValPheLysPheValGlyPheLysPheThrValLysLysAsp---------L 171
Db     4655 AGCGTAAGCGGCAGGCGGAGGTGGAGCTGGCCTCGCGCGTGAAGGCCGAGACCGAGGCGG 4714

Qy      171 ysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyA 191
Db     4715 CGCGCGAGAAGCAGCGGGCCCTGCAGGCCCTGGAGGAGCTGCGGCTGCAGGCGGAGGAGG 4774

Qy      191 laGlu---------------AlaSerValGlyAlaGlyAspHisGlnGluProSerValG 206
Db     4775 CGGAGCGGCGCCTGCGGCAGGCCGAGGTGGAGCGAGCGCGGCAGGTACAGGTGGCCCTGG 4834

Qy      206 luThr------------AlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnS 222
Db     4835 AGACGGCGCAGCGCAGTGCAGAGGCGGAGCTGCAGAGCAAACGCGCCTCCTTCGCCGAGA 4894

Qy      222 erThrGluLysGlnGluGlyThrLeuLysGlnGluGln---SerSerThrGluIleProL 241
Db     4895 AGACGGCACAGCTGGAGCGCTCCCTGCAGGAGGAACACGTGGCTGTGGCACAGCTGCGGG 4954

Qy      241 euGlnAlaGlu---------SerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyG 258
Db     4955 AGGAGGCTGAGCGGCGGGCACAGCAGCAGGCCGAGGCCGAGCGGGCGCGCGAGGAGGCAG 5014

Qy      258 luGluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnS 278
Db     5015 AGCGGGAGCTGGAGCGCTGGCAGCTCAAGGCCAACGAGGCGCTACGGCTGCGGCTGCAGG 5074

Qy      278 erGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysL 298
Db     5075 CGGAGGAGGTGGCGCAGCAGAAGAGCCTG---------------------------GCGC 5107

Qy      298 ysThrSerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluG 318
Db     5108 AGGCCGAGGCTGAGAAGCAGAAGGAGGAGGCGGAGCGCGAGGCGCGGCGGCGCGGCAAGG 5167

Qy      318 lnGluAlaGluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluG 338
Db     5168 CGGAGGAGCAGGCCGTCCGGCAGCGGGAGCTGGCTGAACAAGAG---CTGGAGAAGCAGC 5224

Qy      338 lnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluL 358
Db     5225 GGCAGCTGGCGGAAGGCACCGCGCAGCAGCGCCTGGCCGCGGAGCAGGAGTTGATCCGGC 5284

Qy      358 euProLeuGluAspGlnValGlyAspLeuGluAlaSer---SerGluGluLysCysAlaP 377
Db     5285 TGCGGGCCGAGACGGAGCAGGGGGAGCAGCAGCGGCAGCTGCTGGAGGAGGAGCTGGCCC 5344

Qy      377 roLeu------------------AlaThrGluValPheAspGlu---------------- 385
Db     5345 GGCTGCAGCGTGAGGCGGCTGCAGCCACGCAGAAACGGCAGGAGCTGGAAGCCGAGCTGG 5404

Qy      386 --LysMetGluAlaHisGlnGluValVal---------------------AlaGluValH 398
Db     5405 CCAAGGTGCGGGCCGAGATGGAGGTGCTGCTGGCCAGCAAGGCGAGGGCTGAGGAGGAGT 5464

Qy      398 isValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyG 418
Db     5465 CGCGCTCCACCAGCGAGAAGTCCAAGCAGAGGCTG------------GAGGCCGAGGCCG 5512

Qy      418 lyValValValGluGlyThrGlyGlu---------------SerLeuProProGlu---- 431
Db     5513 GCCGGTTCCGCGAGCTGGCCGAGGAGGCCGCCCGCCTGCGTGCCCTGGCGGAAGAGGCCA 5572

Qy      432 -----------LysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluG 448
Db     5573 AGCGGCAGCGGCAGCTGGCCGAGGAAGACGCGGCGCGGCAGCGGGCCGAG---GCGGAGC 5629

Qy      448 luLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspL 468
Db     5630 GGGTGCTTGCGGAGAAGCTGGCCGCCATCGGC---------------GAGGCCACGCGGC 5674

Qy      468 euSerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluM 488
Db     5675 TCAAGACGGAGGCGGAGATCGCGCTCAAGGAGAAGGAGGCGGAGAACGAGCGCCTGCGGC 5734

Qy      488 etLeuSerSerGlnGlu---------------ArgIleLysValGlnGlySerProLeuL 503
Db     5735 GGCTGGCGGAGGACGAGGCCTTCCAGCGGCGGCGGCTGGAGGAGCAGGCCGCGCAACACA 5794

Qy      503 ysLysLeuPheSerSer------SerGlyLeuLysLysLeuSerGly---------LysL 518
Db     5795 AGGCTGACATCGAGGAGCGCCTGGCCCAGCTGCGCAAGGCATCGGACAGCGAGCTGGAGC 5854

Qy      518 ysGlnLysGlyLysArgGlyGly---------------------GlyGlyAspGluGluP 531
Db     5855 GGCAGAAGGGGCT-GGTGGAGGACACGCTGAGGCAGCGGCGGCAGGTGGAGGAAGAGATC 5913

Qy      531 roGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln------L 549
Db     5914 CTGGC------------GCTGAAGGCGAGCTTCGAGAAGGCGGCCGCTGGCAAGGCGGAG 5961

Qy      549 ysGlyGluSerSerAlaSerSerProGluGlu-ProGluGluThr---ThrCysLeuGlu 567
Db     5962 CTGGAGCTGGAGCTGGGACGCATCCGCAGCAACGCGGAGGACACGCTGCGCAGCAAGGAG 6021

Qy      568 LysGlyProLeuGluAlaProGlnAsp--------------------------------- 576
Db     6022 CAGGCCGAGCTGGAGGCCGCGAGGCAGCGGCAGCTGGCGGCGGAGGAGGAGCGGCGGCGC 6081

Qy      577 GlyGluAlaGluGluGlyThrThrSerAsp---------------------GlyGluLys 589
Db     6082 CGTGAGGCTGAGGAGCGCGTGCAGAAGAGCCTGGCGGCCGAGGAGGAGGCCGCACGGCAG 6141

Qy      590 LysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArg 609
```

```
Db   6142 CGGAAGGCGGCGCTGGAGGAAGTCGAGCGGCTGAAAGCCAACGTGGAGGAGGCGCGGCGC 6201

Qy    610 ValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThr 629

Db   6202 CTGCGGGAGCGAGCGGAGCAGGAGTCGGCGCGGCAGCTGCAGCTGGCCCAGGAGGCCGCC 6261

Qy    630 LeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGlu 649

Db   6262 CAGAAGCGGCTG---------------CAGGCGGAAGAGAAGGCACACGCCTTCGCGGTG 6306

Qy    650 GluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeu 669

Db   6307 CAGCAGAAGGAGCAGGAGCTACAGCAGACGCTGCAGCAGGAGCAGAGC------GTGCTG 6360

Qy    670 IleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGly 689

Db   6361 GACCAGCTGCGCGGCGAGGCGGAGGCGGCCCGGCGGGCGGCTGAGGAGGCGGAGGAGGCC 6420

Qy    690 ---------GlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLys 706

Db   6421 CGGGTGCAGGCGGAGCGTGAGGCGGCGCAGGCCCGGCGGCAGGTGGAAGAGGCCGAGCGG 6480

Qy    707 AspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGln 726

Db   6481 CTGAAGCAGTCG------------------GCAGAGGAGCAGGCACAGGCCCGGGCTCAG 6522

Qy    727 GlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrp 746

Db   6523 GCACAGGCGGCTGCAGAG------------------------------------------ 6540

Qy    747 GluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLys 766

Db   6541 ------------------------------AAGCTGCGCAAGGAGGCCGAGCAAGAGGCG 6570

Qy    767 AlaGluAspSerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSer 786

Db   6571 GCGCGGCGGGCACAGGCGGAGCAGGCGGCCCTGCGGCAGAAGCAGGCAGCTGACGCGGAG 6630

Qy    787 TrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGln 806

Db   6631 ATGGAGAAGCATAAGAAATTCGCCGAGCAGACGCTGCGGCAGAAGGCGCAGGTGGAGCAG 6690

Qy    807 GluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnVal 826

Db   6691 GAGCTGACAACACTGCGG------------CTGCAGCTGGAGGAGACCGAC--------- 6729

Qy    827 ProAlaValValProLeuSerGluTyrAsnAlaValGluArgGlu------------Lys 842

Db   6730 ------------------CACCAGAAGAACCTGCTGGACGAGGAGCTGCAGCGGCTGAAG 6771

Qy    843 MetGlu------------AlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaVal 858

Db   6772 GCGGAGGCCACGGAGGCCGCACGCCAGCGCAGCCAGGTGGAGGAGGAGCTCTTCTCGGTG 6831

Qy    859 TyrValSer---GluGluLeuSerLysThrLeuValHisThrValSerValAlaValIle 877

Db   6832 CGCGTGCAGATGGAGGAGCTGAGCAAG--------------------------------- 6858

Qy    878 AspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSer 897

Db   6859 ------CTCAAGGCACGCATCGAGGCTGAGAACCGCGCACTCATCTTGCGTGACAAGGAC 6912

Qy    898 ValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProVal---GluGluVal 916

Db   6913 AATACGCAGCGCTTCCTGCAGGAGGAGGCTGAGAAGATGAAGCAGGTGGCGGAGGAGGCC 6972

Qy    917 ThrGluLysAspIleIleAlaGluGluThrProValLeuThrGln------------Thr 932

Db   6973 GCGCGGCTGAGTGTGGCGGCCCAAGAGGCTGCGCGACTGCGGCAGCTGGCAGAGGAGGAC 7032

Qy    933 LeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAspPheThrSer 952

Db   7033 CTGGCACAGCAGCGGGCCTTGGCAGAGAAGATGCTCAAGGAGAAGATGCAGGCGGTGCAG 7092

Qy    953 GluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAla 972

Db   7093 GAGGCCACGCGACTCAAGGCTGAGGCGGAACTGCTGCAGCAGCAGAAGGAGCTTGCGCAG 7152

Qy    973 SerGlyAlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerPro 992

Db   7153 GAGCAGGCGCGGCGGCTGCAGGAG------------------------------------ 7176

Qy    993 AspThrThrGluGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGlu 1012

Db   7177 GACAAGGAGCAGATGGCGCAGCAGCTGGCGGAGGAGACGCAGGGCTTCCAGCGGACGCTG 7236

Qy   1013 GluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLys------ 1030

Db   7237 GAGGCCGAGCGGCAGCGGCAGCTGGAGATGAGCGCTGAGGCTGAGCGCCTCAAGCTGCGT 7296

Qy   1031 ------------GluGluSerGlnValProAlaThrGlnThrValGlnArgThrGlySer 1046

Db   7297 GTGGCCGAGATGAGCCGAGCCCAGGCCCGCGCTGAGGAGGACGCCCAGCGCTTCCGGAAG 7356

Qy   1047 LysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys 1066

Db   7357 CAGGCGGAGGAGATCGGTGAGAAGCTGCACCGCACGGAG------CTCGCCACCCAGGAG 7410

Qy   1067 GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGln 1086

Db   7411 AAGGTGACCCTGGTGCAGACACTGGAGATCCAGCGACAGCAGAGTGACCATGATGCCGAG 7470

Qy   1087 GlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrAlaAsp 1106

Db   7471 ------------------------------------CGCCTGCGGGAGGCCATCGCTGAG 7494

Qy   1107 ValAspHis------ValAlaThrCysGlnValIleLysLeuGlnGlnLeu---MetGlu 1123

Db   7495 CTGGAGCGTGAGAAGGAGAAGCTCCAACAGGAGGCCAAACTGCTGCAGCTCAAGTCTGAG 7554

Qy   1124 GlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThr 1143

Db   7555 GAGATGCAGACGGTGCAGCAGGAGCAGCTGCTGCAGGAGACGCAGGCCCTGCAGCAAAGC 7614

Qy   1144 ProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGln 1163

Db   7615 TTCCTCTCTGAAAAGGACAGCCTG------CTACAGCGGGAGCGCTTCATCGAGCAGGAG 7668

Qy   1164 AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluAlaAlaThr 1183

Db   7669 AAGGCCAAGCTG------------------GAGCAGCTCTTCCAGGACGAGGTGGCCAAG 7710

Qy   1184 AlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGly 1203

Db   7711 GCACAGCAG-------------------------------CTGCGTGAGGAG------ 7731

Qy   1204 GluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaVal 1223

Db   7732 CAGCAGCGGCAGCAGCAGCAGATGGAGCAGGAACGGCAGCGGCTGGTGGCCAGCATGGAG 7791

Qy   1224 ProValLeuAlaLysThrGluValGlyGlnGluGly-----------------GluVal 1237

Db   7792 GAGGCGCGGCGGCGGCAGCATGAGGCCGAGGAGGGCGTGCGGCGCAAGCAGGAGGAGCTG 7851

Qy   1238 AspTrpLeuAspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyPro 1257

Db   7852 CAGCAGCTGGAGCAGCAGCGG---CGGCAGCAGGAGGAGCTGCTGGCTGAGGAGAACCAG 7908

Qy   1258 AsnSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCys 1277

Db   7909 AGGCTGCGTGAGCAGCTGCAGCTCCTGGAGGAGCAGCACCGGGCCGCGCTGGCGCACTCA 7968

Qy   1278 GlnGluLysGluSerThrGluValGlnSerLeuSer------------------------ 1289

Db   7969 GAGGAGGTCACTGCCTCGCAGGTGGCTGCCACAAAGACCCTGCCCAATGGCCGGGATGCA 8028
```

```
Qy   1290 ------------LeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGlu 1305
Db   8029 CTTGATGGCCCCGCGGCAGAGGCAGAGCCGGAGCACAGCTTCGATGGCCTGCGGCGGAAG 8088

Qy   1306 ThrLysProGluGlnValSerGluGluGlyGluGlnGluThr------------------ 1319
Db   8089 GTGTCAGCTCAGAGGCTGCAGGAGGCCGGCAT-CCTGAGTGCGGAGGAGCTGCAGCGGTT 8147

Qy   1320 ---AlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetPro 1338
Db   8148 GGCGCAGGGCCACACCACGGTGGA------------------------------------ 8171

Qy   1339 SerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGln 1358
Db   8172 ---------CGAGCTCGCACGGCGGGAAGACGTGCGCCACTACCTGCAGGGCCGCAGCAG 8222

Qy   1359 AspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThr 1378
Db   8223 TATCGCAGGGCTGTTGCTGAAGGCCAC-----------------------CAATGAGAA 8258

Qy   1379 AlaGluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerPro 1398
Db   8259 GCTGAGTGTTTACGCCGCCCTGCAGAGGCAGCTGCTGAGTCCCGGCACGGCCCTCATCCT 8318

Qy   1399 Glu--------------------------------------------------------- 1399
Db   8319 GCTGGAGGCGCAGGCGGCCTCAGGCTTCCTGCTGGACCCTGTGCGGAACCGGCGGCTGAC 8378

Qy   1400 ------------------------CysValGlyAlaHisLeuLeuProAlaGluLysSer 1411
Db   8379 CGTCAACGAGGCTGTGAAGGAGGGTGTGGTGGGCCCCGAGCTGCACCACAAGCTGCTGTC 8438

Qy   1412 SerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyPro 1431
Db   8439 GGCCGAGCGCGCCGTCACTGGCTACAAGGACCCCTACACTGGCCAGCAGATCTCTCTCTT 8498

Qy   1432 GluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHis 1451
Db   8499 CCAAGCCATGCAGAAGGG------------------CCTCATCGTCCGGGAGCACGGCAT 8540

Qy   1452 ProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLys 1471
Db   8541 CCG--------------------------------------------------------- 8543

Qy   1472 ProAspAlaGlyProAspAlaAspGlyLys 1481
Db   8544 CCTGCTGGAGGCCCAGATCGCCACGGGCGG 8573
```

RESULT 12
US-09-724-676-7257
; Sequence 7257, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7257
;   LENGTH: 8161
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-7257

Alignment Scores:

Pred. No.:	1.98e-09	Length:	8161
Score:	415.00	Matches:	316
Percent Similarity:	38.17%	Conservative:	305
Best Local Similarity:	19.42%	Mismatches:	669
Query Match:	5.14%	Indels:	337
DB:	5	Gaps:	60

US-09-902-432-4 (1-1596) x US-09-724-676-7257 (1-8161)

```
Qy     26 LeuSerGly-----------HisGlyProAlaAlaGluAlaSerGlyAlaAla------ 39
Db   2376 CTCTCAGGTGAAGGAACTTAGCATGGTAACAGAATTGAGGGCTCAGGTAAAGCAACTGGA 2435

Qy     40 GlyAspPro-AlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSe 59
Db   2436 AATGAACCTTGCAGAAGCAGAAAGGCAAAGAAGACTTGATTATGAAAGC---CAAACTGC 2492

Qy     59 rSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGl 79
Db   2493 CCATGACAACCTGCTCACTGAACAG------ATCCATAGTCTCAGCATAGAAGCCAAATC 2546

Qy     79 yGlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGl 99
Db   2547 TAAAGATGTGAAAATT---------GAAGTTTTACAGAATGAACTGGATGATGTGCAG-- 2595

Qy     99 uLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAs 119
Db   2596 ------------CTTCAGTTTTCTGAGCAGAGTACCCTGATAAGAAGCCTGCAAAGCCA 2642

Qy    119 pGlyGlnGluGluThrSerGluIleIle---------GluGlnIleProAlaSerGluAs 136
Db   2643 GCTGCAAAATAAGGAAAGTGAAGTGCTTGAGGGGGCAGAACGTGTAAGGCATATCTCAAG 2702

Qy    136 nAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLy 156
Db   2703 TAAAGTGGAAGAACTGTCCCAGGCTCTTTCACAGAAGGAACTTGAAATA----------- 2751

Qy    156 sValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAs 176
Db   2752 ------------------------------ACAAAAATGGAT---------------- 2763

Qy    176 pThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGl 196
Db   2764 -------CAGCTCTTACTAGAGAAAAAGAGAGATGTGGAAACCCTCCAACAAACCATCGA 2816

Qy    196 yAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGl 216
Db   2817 GGAGAAGGATCAACAAGTGACAGAAATCAGCTTTAGTATGACTGAGAAAATGGTTCAGCT 2876

Qy    216 uSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSe 236
Db   2877 TAATGAAGAGAAGTTTTCTCTTGGGGTTGAAATTAAGACTCTTAAAGAA----------- 2925

Qy    236 rThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGl 256
Db   2926 ----CAGCTAAATTTATTATCCAGAGCTGAGGAAGCAAAAAAAGAGCAGGTGGAAGAAGA 2981

Qy    256 uGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProVa 276
Db   2982 TAATGAAGTTTCTTCTGGCCTTAAACAAAATTATGATGAGATGAGCCCAGCAGGACAAAT 3041

Qy    276 lAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpAr 296
Db   3042 AAGTAAGGAAGAA--------------------CTTCAGCATGAATTTGACCTTCTGAA 3080

Qy    296 gLysLysThrSerPheLysLysSerLys----------------------GluAspAs 308
Db   3081 GAAAGAAAATGAGCAGAGAAAGAGAAAGCTCCAGGCAGCTCTTATTAACAGAAAGGAGCT 3140

Qy    308 pLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLy 328
Db   3141 TCTGCAAAGAGTCAGTAGATTGGAAGAAGAATTAGCCAACTTGAAAGATGAATCTAAGAA 3200

Qy    328 sGluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaAr 348
Db   3201 AGAAATCCCACTCAGTGAGACTGAGAGGGGAGAAGTGGAAGAAGATAAAGAAAACAAAGA 3260

Qy    348 gLeuSer---------------AlaAspTyrGluLysValGluLeuProLeuGluAspGl 363
```

```
Db   3261 ATACTCAGAAAAATGTGTGACTTCTAAGTGCCAAGAAATAGAAATTTATTTAAAACAGAC 3320

Qy    363 nValGlyAspLeuGluAla--------------------SerSerGluGluLysCysAl 376
Db   3321 AATATCTGAGAAAGAAGTGGAACTACAGCATATAAGGAAGGATTTGGAAGAAAAG----- 3375

Qy    376 aProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGl 396
Db   3376 ----CTGGCAGCTGAA------------------GAGCAATTCCAGGCTCTGGTCAAACA 3413

Qy    396 uValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGl 416
Db   3414 GATGAATCAGACCTTGCAAGATAAAACAAACCAA-------------------------- 3447

Qy    416 uGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluPr 436
Db   3448 ----------------------------ATAGATTTGCTCCAAGCAGAAATCAGTGAAAA 3479

Qy    436 oGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCy 456
Db   3480 CCAAGCAATTATCCAGAAG---------------------TTAATCACAAGTAACACGGA 3518

Qy    456 sValSerGlyGlyAspHisThrGlnLeuThrAsp------------LeuSerProGluGl 472
Db   3519 TGCAAGTGATGGGGACTCCGTAGCACTTGTAAAGGAAACAGTGGTGATAAGTCCACCTTG 3578

Qy    472 uLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGl 492
Db   3579 TACAGGTAGTAGTGAACACTGGAAACCAGAACTAGAAGAAAAGATACTGGCCCTTGAAAA 3638

Qy    492 nGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLy 512
Db   3639 AGAAAAGGAGCAACTTCAAAAGAAGCTACAGGAAGCCTTAACCTCCCGCAAGGCAATTCT 3698

Qy    512 sLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGl 532
Db   3699 TAAAAAGGCACAGGAGAAAGAAAGACATCTCAGG---------------GAGGAGCTAAA 3743

Qy    532 yGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSe 552
Db   3744 GCAACAGAAAGATGACTATAATCGCTTGCAAGAACAGTTTGATGAGCAAAGCAAGGAAAA 3803

Qy    552 r------------------------------------------------------SerAl 554
Db   3804 TGAGAATATTGGAGACCAGCTAAGGCAACTCCAGATTCAAGTAAGGGAATCCATAGACGG 3863

Qy    554 aSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaPr 574
Db   3864 AAAACTCCCAAGCACAGACCAGCAGGAATCGTGTTCTTCCACTCCAGGTTTAGAAGAACC 3923

Qy    574 oGlnAspGlyGluAlaGluGluGlyThrThr----------------------------- 584
Db   3924 TTTATTCAAAGCCACAGAACAGCATCACACTCAACCTGTTTTAGAGTCCAACTTGTGCCC 3983

Qy    585 ----------------SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSe 599
Db   3984 AGACTGGCCTTCTCATTCTGAAGATGCGAGTGCTCTGCAGGGCGGAACTTCTGTTGCCCA 4043

Qy    599 rPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGl 619
Db   4044 GATTAAGGCCCAGCTG---------------------------------------AAGGA 4064

Qy    619 uGluGluLeuGluLysValLys---SerAlaThrLeuSerSerThrAspSerThrValSe 638
Db   4065 AATAGAGGCTGAGAAAGTAGAGTTAGAATTGAAAGTTAGTTCTACAACAAGTGAGCTTAC 4124

Qy    638 rGluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysAr 658
Db   4125 TAAAAAATCAGAAGAGGTATTTCAGTTACAAGAGCAGATAAATAAACAGGGTTTAGAAAT 4184

Qy    658 gArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysAr 678
Db   4185 CGAGAGTCTAAAGACAGTATCCCATGAAGCTGAAGTCCATGCCGAAAGCCTGCAGCAGAA 4244

Qy    678 gAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSe 698
Db   4245 ATTG------------------------GAAAGCAGCCAACTACAAATTGCTGGCCTAGA 4280

Qy    698 rHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSe 718
Db   4281 ACATCTAAGAGAATTGCAACCTAAACTGGATGAACTG---------CAAAAACTCATAAG 4331

Qy    718 rThrGlnGluGlnAspGlnAlaGlnGlySer---------SerSerProGluProAlaGl 735
Db   4332 CAAAAAGGAAGAAGACGTTAGCTACCTTTCTGGACAACTTAGTGAGAAAGAAGCAGCTCT 4391

Qy    735 ySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPr 755
Db   4392 CACTAAAATACAGACAGAGATAATAGAACAAGAAGATTTAATTAAGGCTCTGCATACA-- 4449

Qy    755 oArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLe 775
Db   4450 -------------CAGCTAGAAATGCAAGCCAAAGAGCATGATGAGAGGATAAAGCAGCT 4496

Qy    775 uSerThrGluIle---------------------GluProSerArgGluGluSerTrpVa 788
Db   4497 ACAGGTGGAACTTTGTGAAATGAAGCAAAAACCAGAAGAGATTGGAGAAGAAAGT----- 4551

Qy    788 lSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGl 808
Db   4552 -------------------------AGAGCAAAGCAACAAATACAAAGGAAACTGCAAGC 4586

Qy    808 nAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAl 828
Db   4587 TGCCCTTATTTCCCGAAAA---------GAAGCACTAAAAGAAAACAAAAGTCTCCAAGA 4637

Qy    828 aValValProLeuSerGluTyrAsnAlaValGluArg----------------------- 840
Db   4638 GGAATTGTCTTTGGCCAGA---GGTACCATTGAACGTCTCACCAAGTCTCTGGCAGATGT 4694

Qy    841 -GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVa 860
Db   4695 GGAAAGCCAAGTTTCTGCTCAAAATAAAGAAAAAGATACGGTCTTAGGAAGGTTAGCTCT 4754

Qy    860 lSerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyTh 880
Db   4755 TCTTCAAGAAGAAAGAGACAAACTCATTACAGAAATGGACAGGTCTTTATTGGAAAATCA 4814

Qy    880 rArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGl 900
Db   4815 GAGTCTCAGCAGCTCCTGTGAA------AGTCTAAAACTAGCTCTAGAGGGTCTTACTGA 4868

Qy    900 uProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAs 920
Db   4869 AGACAAGGAAAAGTTAGTGAAGGAAATT---------------GAATCTTTGAAATCTTC 4913

Qy    920 pIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHi 940
Db   4914 TAAGATTGCAGAAAGTACT------------------------GAGTGGCAAGAGAAACA 4949

Qy    940 sAspAsp---------------------MetValThrSerGluValAspPheThrSerGl 953
Db   4950 CAAGGAGCTACAAAAAGAGTATGAAATTCTTCTGCAGTCCTATGAGAATGTTAGTAATGA 5009

Qy    953 uAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGlu-------------- 968
Db   5010 AGCAGAAAGGATTCAGCATGTGGTGGAAGCTGTGAGGCAAGAGAAACAAGAACTGTATGG 5069

Qy    969 ----------ValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaVa 985
Db   5070 CAAGTTAAGAAGCACAGAGGCAAACAAGAAGGAGACAGAAAAGCAGTTGCAGGAAGCTGA 5129

Qy    985 lSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGl 1005
Db   5130 GCAAGAAATGGAGGAAATGAAAGAAAAGATGAGAAAGTTT------GCTAAATCTAAACA 5183
```

```
Qy   1005 uSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaVa 1025
Db   5184 GCAGAAAATCCTAGAGCTGGAAGAAGAGAATGACCGGCTTAGGGCAGAGGTGCACCCTGC 5243

Qy   1025 lAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrGl 1045
Db   5244 AGGAGATACAGCTAAAGAG---------------TGTATGGAAACACTTCTTTCTTCCAA 5288

Qy   1045 ySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGl 1065
Db   5289 TGCCAGCATGAAGGAAGAACTTGAAAGGGTCAAAATGGAGTATGAAACCCTTTCTAAGAA 5348

Qy   1065 u---------------LysGluLysAspValMetProLysGlyProValGlnGluAlaGl 1080
Db   5349 GTTTCAGTCTTTAATGTCTGAGAAAGACTCTCTAAGTGAA---GAGGTTCAAGAT----- 5400

Qy   1080 yAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVa 1100
Db   5401 -TTAAAGCATCAGATAGAAGGTAATGTATCTAAACAAGCTAACCTAGAGGCCACCGAGAA 5459

Qy   1100 lProGluValThrAlaAspValAspHisValAlaThrCysGlnVal---------IleLy 1117
Db   5460 ACATGATAACCAAACGAATGTCACTGAAGAGGGAACACAGTCTATACCAGGTGAGACTGA 5519

Qy   1117 sLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSe 1137
Db   5520 AGAGCAAGACTCTCTGAGTATGAGCACAAGACCTACATGTTCAGAATCGGTTCCATCAGC 5579

Qy   1137 rGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAs 1157
Db   5580 GAAGAGT------GCCAACCCTGCTGTAAGTAAGGATTTCAGCTCACATGATGAAATTAA 5633

Qy   1157 p---------GluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSe 1174
Db   5634 TAACTACCTACAGCAGATTGATCAGCTCAAAGAAAGAATTGCTGGATTAGAGGAGGAGAA 5693

Qy   1174 rGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAs 1194
Db   5694 GCAGAAAAACAAGGAATTTAGCCAGACTTTAGAAAATGAGAAA---AATACCTTACTGAG 5750

Qy   1194 nAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProTh 1214
Db   5751 TCAGATATCAACAAAGGAT------GGTGAA------------------CTAAAAATGCT 5786

Qy   1214 rGlnGlnGluLeuThr---------------------------AlaAlaAlaValProVa 1225
Db   5787 TCAGGAGGAAGTAACCAAAATGAACCTGTTAAATCAGCAAATCCAAGAAGAACTCTCCAG 5846

Qy   1225 lLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysVa 1245
Db   5847 AGTTACCAAACTAAAGGAGACAGCAGAAGAAGAGAAAGATGATTTGGAAGAGAGGCTTAT 5906

Qy   1245 lLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspVa 1265
Db   5907 GAATCAATTAGCAGAACTT---------AATGGAAGCATTGGGAATTACTGTCAGGATGT 5957

Qy   1265 lThr------------------------TyrAspSerGluValMetGlyValAlaGlyCy 1277
Db   5958 TACAGATGCCCAAATAAAAAATGAGCTATTGGAATCTGAAATGAAGAACCTTAAAAAGTG 6017

Qy   1277 sGlnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluTh 1297
Db   6018 TGTGAGTGAATTGGAAGAAGAAAAGCAGCAGTTAGTCAAGGAAAAAACTAAGGTGGAATC 6077

Qy   1297 rAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGl 1317
Db   6078 AGAAATACGAAAGGAATATTTGGAGAAAATACAAGGTGCTCAGAAAGAACCCGGAAATAA 6137

Qy   1317 nGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMe 1337
Db   6138 A--------------------------AGCCATGCAAAGGAACTTCAGGAACTGTTAAA 6170

Qy   1337 tProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAs 1357
Db   6171 AGAAAAACAACAAGAAGTAAAGCAGCTA------------CAGAAGGACTGCATCAGGTA 6218

Qy   1357 pGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGl 1377
Db   6219 TCAAGAGAAAATTAGTGCTCTGGAGAGAACTGTTAAAGCTCTAGAATTTGTTCAAACTGA 6278

Qy   1377 nThrAlaGluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSe 1397
Db   6279 ATCTCAAAAAGATTTGGAAATAACCAAAGAAAAT------CTGGCTCAAGCAGTTGAACA 6332

Qy   1397 rProGluCysValGlyAlaHis------------------LeuLeuProAlaGluLysSe 1411
Db   6333 CCGCAAAAAGGCACAAGCAGAATTAGCTAGCTTCAAAGTCCTGCTAGATGACACTCAAAG 6392

Qy   1411 rSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyPr 1431
Db   6393 TGAAGCAGCA------------------AGGGTCCTAGCAGACAATCTCAAGTTGAAAAA 6434

Qy   1431 oGluSerGlnAla------GluSerIleProIleIleValThrProAlaProGluSerTh 1449
Db   6435 GGAACTTCAGTCAAATAAAGAATCAGTTAAAAGCCAGATGAAACAAAAGGATGAA----- 6489

Qy   1449 rLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGl 1469
Db   6490 ----------GATCTTGAGCGAAGACTGGAACAGGCAGAAGAGAAGCACCTGAAAGAGAA 6539

Qy   1469 uAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysVa 1489
Db   6540 GAAGAATATGCAAGAGAAACTGGATGCTTTGCGCAGAGAAAAAGTCCACTTGGAAGAGAC 6599

Qy   1489 lLeu-----------------------LysAlaGluProGluIleLeuGluLeuGluSe 1501
Db   6600 AATTGGAGAGATTCAGGTTACTTTGAACAAGAAAGACAAGGAAGTTCAGCAACTTCAGGA 6659

Qy   1501 rLysSerAsnLysIleVal 1507
Db   6660 AAACTTGGACAGTACTGTG 6678
```

RESULT 13
US-09-724-676A-7257
; Sequence 7257, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7257
;   LENGTH: 8161
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676A-7257

Alignment Scores:

Pred. No.:	1.98e-09	Length:	8161
Score:	415.00	Matches:	316
Percent Similarity:	38.17%	Conservative:	305
Best Local Similarity:	19.42%	Mismatches:	669
Query Match:	5.14%	Indels:	337
DB:	5	Gaps:	60

US-09-902-432-4 (1-1596) x US-09-724-676A-7257 (1-8161)

```
Qy     26 LeuSerGly-----------HisGlyProAlaAlaGluAlaSerGlyAlaAla------ 39
Db   2376 CTCTCAGGTGAAGGAACTTAGCATGGTAACAGAATTGAGGGCTCAGGTAAAGCAACTGGA 2435

Qy     40 GlyAspPro-AlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSe 59
Db   2436 AATGAACCTTGCAGAAGCAGAAAGGCAAAGAAGACTTGATTATGAAAGC---CAAACTGC 2492
```

```
Qy     59 rSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGl 79
Db   2493 CCATGACAACCTGCTCACTGAACAG------ATCCATAGTCTCAGCATAGAAGCCAAATC 2546

Qy     79 yGlnGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGl 99
Db   2547 TAAAGATGTGAAAATT---------GAAGTTTTACAGAATGAACTGGATGATGTGCAG-- 2595

Qy     99 uLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAs 119
Db   2596 -------------CTTCAGTTTTCTGAGCAGAGTACCCTGATAAGAAGCCTGCAAAGCCA 2642

Qy    119 pGlyGlnGluGluThrSerGluIleIle---------GluGlnIleProAlaSerGluAs 136
Db   2643 GCTGCAAAATAAGGAAAGTGAAGTGCTTGAGGGGGCAGAACGTGTAAGGCATATCTCAAG 2702

Qy    136 nAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLy 156
Db   2703 TAAAGTGGAAGAACTGTCCCAGGCTCTTTCACAGAAGGAACTTGAAATA----------- 2751

Qy    156 sValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAs 176
Db   2752 ------------------------------ACAAAAATGGAT----------------- 2763

Qy    176 pThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGl 196
Db   2764 -------CAGCTCTTACTAGAGAAAAAGAGAGATGTGGAAACCCTCCAACAAACCATCGA 2816

Qy    196 yAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGl 216
Db   2817 GGAGAAGGATCAACAAGTGACAGAAATCAGCTTTAGTATGACTGAGAAAATGGTTCAGCT 2876

Qy    216 uSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSe 236
Db   2877 TAATGAAGAGAAGTTTTCTCTTGGGGTTGAAATTAAGACTCTTAAAGAA----------- 2925

Qy    236 rThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGl 256
Db   2926 ----CAGCTAAATTTATTATCCAGAGCTGAGGAAGCAAAAAAAGAGCAGGTGGAAGAAGA 2981

Qy    256 uGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProVa 276
Db   2982 TAATGAAGTTTCTTCTGGCCTTAAACAAAATTATGATGAGATGAGCCCAGCAGGACAAAT 3041

Qy    276 lAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpAr 296
Db   3042 AAGTAAGGAAGAA--------------------CTTCAGCATGAATTTGACCTTCTGAA 3080

Qy    296 gLysLysThrSerPheLysLysSerLys-----------------------GluAspAs 308
Db   3081 GAAAGAAAATGAGCAGAGAAAGAGAAAGCTCCAGGCAGCTCTTATTAACAGAAAGGAGCT 3140

Qy    308 pLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLy 328
Db   3141 TCTGCAAAGAGTCAGTAGATTGGAAGAAGAATTAGCCAACTTGAAAGATGAATCTAAGAA 3200

Qy    328 sGluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaAr 348
Db   3201 AGAAATCCCACTCAGTGAGACTGAGAGGGGAGAAGTGGAAGAAGATAAAGAAAACAAAGA 3260

Qy    348 gLeuSer---------------AlaAspTyrGluLysValGluLeuProLeuGluAspGl 363
Db   3261 ATACTCAGAAAAAATGTGTGACTTCTAAGTGCCAAGAAATAGAAATTTATTTAAAACAGAC 3320

Qy    363 nValGlyAspLeuGluAla--------------------SerSerGluGluLysCysAl 376
Db   3321 AATATCTGAGAAAGAAGTGGAACTACAGCATATAAGGAAGGATTTGGAAGAAAAG----- 3375

Qy    376 aProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGl 396
Db   3376 ----CTGGCAGCTGAA-----------------GAGCAATTCCAGGCTCTGGTCAAACA 3413

Qy    396 uValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGl 416
Db   3414 GATGAATCAGACCTTGCAAGATAAAACAAACCAA------------------------- 3447

Qy    416 uGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluPr 436
Db   3448 ---------------------------ATAGATTTGCTCCAAGCAGAAATCAGTGAAAA 3479

Qy    436 oGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCy 456
Db   3480 CCAAGCAATTATCCAGAAG--------------------TTAATCACAAGTAACACGGA 3518

Qy    456 sValSerGlyGlyAspHisThrGlnLeuThrAsp------------LeuSerProGluGl 472
Db   3519 TGCAAGTGATGGGGACTCCGTAGCACTTGTAAAGGAAACAGTGGTGATAAGTCCACCTTG 3578

Qy    472 uLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGl 492
Db   3579 TACAGGTAGTAGTGAACACTGGAAACCAGAACTAGAAGAAAAGATACTGGCCCTTGAAAA 3638

Qy    492 nGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLy 512
Db   3639 AGAAAAGGAGCAACTTCAAAAGAAGCTACAGGAAGCCTTAACCTCCCGCAAGGCAATTCT 3698

Qy    512 sLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGl 532
Db   3699 TAAAAAGGCACAGGAGAAAGAAAGACATCTCAGG---------------GAGGAGCTAAA 3743

Qy    532 yGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSe 552
Db   3744 GCAACAGAAAGATGACTATAATCGCTTGCAAGAACAGTTTGATGAGCAAAGCAAGGAAAA 3803

Qy    552 r----------------------------------------------------SerAl 554
Db   3804 TGAGAATATTGGAGACCAGCTAAGGCAACTCCAGATTCAAGTAAGGGAATCCATAGACGG 3863

Qy    554 aSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaPr 574
Db   3864 AAAACTCCCAAGCACAGACCAGCAGGAATCGTGTTCTTCCACTCCAGGTTTAGAAGAACC 3923

Qy    574 oGlnAspGlyGluAlaGluGluGlyThrThr---------------------------- 584
Db   3924 TTTATTCAAAGCCACAGAACAGCATCACACTCAACCTGTTTTAGAGTCCAACTTGTGCCC 3983

Qy    585 ----------------SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSe 599
Db   3984 AGACTGGCCTTCTCATTCTGAAGATGCGAGTGCTCTGCAGGGCGGAACTTCTGTTGCCCA 4043

Qy    599 rPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGl 619
Db   4044 GATTAAGGCCCAGCTG------------------------------------------AAGGA 4064

Qy    619 uGluGluLeuGluLysValLys---SerAlaThrLeuSerSerThrAspSerThrValSe 638
Db   4065 AATAGAGGCTGAGAAAGTAGAGTTAGAATTGAAAGTTAGTTCTACAACAAGTGAGCTTAC 4124

Qy    638 rGluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysAr 658
Db   4125 TAAAAAATCAGAAGAGGTATTTCAGTTACAAGAGCAGATAAATAAACAGGGTTTAGAAAT 4184

Qy    658 gArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysAr 678
Db   4185 CGAGAGTCTAAAGACAGTATCCCATGAAGCTGAAGTCCATGCCGAAAGCCTGCAGCAGAA 4244

Qy    678 gAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSe 698
Db   4245 ATTG-----------------------GAAAGCAGCCAACTACAAATTGCTGGCCTAGA 4280

Qy    698 rHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSe 718
Db   4281 ACATCTAAGAGAATTGCAACCTAAACTGGATGAACTG---------CAAAAACTCATAAG 4331

Qy    718 rThrGlnGluGlnAspGlnAlaGlnGlySer---------SerSerProGluProAlaGl 735
```

```
                |   :::|||:::|||   :::        |||         |||      |||   |||
Db     4332 CAAAAAGGAAGAAGACGTTAGCTACCTTTCTGGACAACTTAGTGAGAAAGAAGCAGCTCT 4391

Qy      735 ySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPr 755
             :::      :::  |||  :::         :::      |||  |||   |||
Db     4392 CACTAAAATACAGACAGAGATAATAGAACAAGAAGATTTAATTAAGGCTCTGCATACA-- 4449

Qy      755 oArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLe 775
                           :::    :::::::::    |||      |||      ::::::|||||
Db     4450 -------------CAGCTAGAAATGCAAGCCAAAGAGCATGATGAGAGGATAAAGCAGCT 4496

Qy      775 uSerThrGluIle--------------------GluProSerArgGluGluSerTrpVa 788
            |   |||:::                       |||          |||||||||
Db     4497 ACAGGTGGAACTTTGTGAAATGAAGCAAAAACCAGAAGAGATTGGAGAAGAAAGT----- 4551

Qy      788 lSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGl 808
                                         |||   |||::::::          |||   :::
Db     4552 ------------------------AGAGCAAAGCAACAAATACAAAGGAAACTGCAAGC 4586

Qy      808 nAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAl 828
             |||   :::                 |||        :::::::::  ::::::
Db     4587 TGCCCTTATTTCCCGAAAA---------GAAGCACTAAAAGAAAACAAAAGTCTCCAAGA 4637

Qy      828 aValValProLeuSerGluTyrAsnAlaValGluArg---------------------- 840
                :::   |||:::            :::||||||
Db     4638 GGAATTGTCTTTGGCCAGA---GGTACCATTGAACGTCTCACCAAGTCTCTGGCAGATGT 4694

Qy      841 -GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVa 860
             |||      :::         |||  |||      :::||||||    :::   ::
Db     4695 GGAAAGCCAAGTTTCTGCTCAAAATAAAGAAAAAGATACGGTCTTAGGAAGGTTAGCTCT 4754

Qy      860 lSerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyTh 880
            :    :::|||           |||:::         :::      ::::::::::::
Db     4755 TCTTCAAGAAGAAAGAGACAAACTCATTACAGAAATGGACAGGTCTTTATTGGAAAATCA 4814

Qy      880 rArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGl 900
                           :::|||   |||      |||                :::|||||
Db     4815 GAGTCTCAGCAGCTCCTGTGAA----AGTCTAAAACTAGCTCTAGAGGGTCTTACTGA 4868

Qy      900 uProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAs 920
            |       |||          |||                     |||     :::
Db     4869 AGACAAGGAAAAGTTAGTGAAGGAAATT--------------GAATCTTTGAAATCTTC 4913

Qy      920 pIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHi 940
               |||||||||   |||                            |||   ::::::  ||
Db     4914 TAAGATTGCAGAAAGTACT-----------------------GAGTGGCAAGAGAAACA 4949

Qy      940 sAspAsp--------------------MetValThrSerGluValAspPheThrSerGl 953
            |   :::                     ::::::   |||      :::    ::::::||
Db     4950 CAAGGAGCTACAAAAAGAGTATGAAATTCTTCTGCAGTCCTATGAGAATGTTAGTAATGA 5009

Qy      953 uAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGlu------------- 968
             ||||                        ||||||:::|||    |||:::
Db     5010 AGCAGAAAGGATTCAGCATGTGGTGGAAGCTGTGAGGCAAGAGAAACAAGAACTGTATGG 5069

Qy      969 ---------ValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaVa 985
                       |||||||||:::         |||               :::      |||
Db     5070 CAAGTTAAGAAGCACAGAGGCAAACAAGAAGGAGACAGAAAAGCAGTTGCAGGAAGCTGA 5129

Qy      985 lSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGl 1005
               ::::::   :::      :::        :::           :::        ::
Db     5130 GCAAGAAATGGAGGAAATGAAAGAAAAGATGAGAAAGTTT------GCTAAATCTAAACA 5183

Qy     1005 uSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaVa 1025
            :       :::|||:::   |||||||||           :::    :::|||   :::
Db     5184 GCAGAAAATCCTAGAGCTGGAAGAAGAGAATGACCGGCTTAGGGCAGAGGTGCACCCTGC 5243

Qy     1025 lAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrGl 1045
                |||      ||||||                         :::|||:::      :::
Db     5244 AGGAGATACAGCTAAAGAG----------------TGTATGGAAACACTTCTTTCTTCCAA 5288

Qy     1045 ySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGl 1065
             :::           |||::::::|||   |||:::   :::   |||   |||:::    ::
Db     5289 TGCCAGCATGAAGGAAGAACTTGAAAGGGTCAAAATGGAGTATGAAACCCTTTCTAAGAA 5348

Qy     1065 u---------------LysGluLysAspValMetProLysGlyProValGlnGluAlaGl 1080
            :                   |||||||||   :::   :::         ||||||:::
Db     5349 GTTTCAGTCTTTAATGTCTGAGAAAGACTCTCTAAGTGAA---GAGGTTCAAGAT----- 5400

Qy     1080 yAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVa 1100
                 :::|||        :::|||:::    :::      ||||||      |||:::     |||
Db     5401 -TTAAAGCATCAGATAGAAGGTAATGTATCTAAACAAGCTAACCTAGAGGCCACCGAGAA 5459

Qy     1100 lProGluValThrAlaAspValAspHisValAlaThrCysGlnVal---------IleLy 1117
                :::          :::|||            |||      :::             ::
Db     5460 ACATGATAACCAAACGAATGTCACTGAAGAGGGAACACAGTCTATACCAGGTGAGACTGA 5519

Qy     1117 sLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSe 1137
            :    |||         :::        :::    |||      ||||||::::::
Db     5520 AGAGCAAGACTCTCTGAGTATGAGCACAAGACCTACATGTTCAGAATCGGTTCCATCAGC 5579

Qy     1137 rGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAs 1157
            ::::::      :::    |||                |||     :::          :::  ::
Db     5580 GAAGAGT------GCCAACCCTGCTGTAAGTAAGGATTTCAGCTCACATGATGAAATTAA 5633

Qy     1157 p---------GluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSe 1174
            :             :::   ||||||                :::                :::
Db     5634 TAACTACCTACAGCAGATTGATCAGCTCAAAGAAAGAATTGCTGGATTAGAGGAGGAGAA 5693

Qy     1174 rGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAs 1194
              |||         :::|||   :::   |||    :::     |||:::     :::||||||    ::
Db     5694 GCAGAAAAACAAGGAATTTAGCCAGACTTTAGAAAATGAGAAA---AATACCTTACTGAG 5750

Qy     1194 nAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProTh 1214
            :      :::        ::::::      ||||||                |||:::
Db     5751 TCAGATATCAACAAAGGAT------GGTGAA------------------CTAAAAATGCT 5786

Qy     1214 rGlnGlnGluLeuThr---------------------------AlaAlaAlaValProVa 1225
              |||:::|||:::|||                                       :::
Db     5787 TCAGGAGGAAGTAACCAAAATGAACCTGTTAAATCAGCAAATCCAAGAAGAACTCTCCAG 5846

Qy     1225 lLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysVa 1245
              :::    |||   :::             |||   |||   |||   |||:::          ::
Db     5847 AGTTACCAAACTAAAGGAGACAGCAGAAGAAGAGAAAGATGATTTGGAAGAGAGGCTTAT 5906

Qy     1245 lLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspVa 1265
              :    :::          |||:::                :::|||                   |||||
Db     5907 GAATCAATTAGCAGAACTT---------AATGGAAGCATTGGGAATTACTGTCAGGATGT 5957

Qy     1265 lThr-----------------------TyrAspSerGluValMetGlyValAlaGlyCy 1277
             ||||                              :::||||||:::           :::       ||
Db     5958 TACAGATGCCCAAATAAAAAATGAGCTATTGGAATCTGAAATGAAGAACCTTAAAAAGTG 6017

Qy     1277 sGlnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluTh 1297
            |       :::              |||   |||   |||   |||:::   ::::::|||::
Db     6018 TGTGAGTGAATTGGAAGAAGAAAAGCAGCAGTTAGTCAAGGAAAAAACTAAGGTGGAATC 6077

Qy     1297 rAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGl 1317
            :::::::      |||||            |||                    |||     |||     ::
Db     6078 AGAAATACGAAAGGAATATTTGGAGAAAATACAAGGTGCTCAGAAAGAACCCGGAAATAA 6137

Qy     1317 nGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMe 1337
            :                                ::::::  |||   :::       |||
Db     6138 A--------------------------AGCCATGCAAAGGAACTTCAGGAACTGTTAAA 6170

Qy     1337 tProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAs 1357
                      :::          |||   |||                         :::
Db     6171 AGAAAAACAACAAGAAGTAAAGCAGCTA------------CAGAAGGACTGCATCAGGTA 6218

Qy     1357 pGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGl 1377
               |||:::|||            :::|||          |||::::::|||:::          |||::
Db     6219 TCAAGAGAAAATTAGTGCTCTGGAGAGAACTGTTAAAGCTCTAGAATTTGTTCAAACTGA 6278

Qy     1377 nThrAlaGluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSe 1397
              ::::   :::    :::|||             |||        :::::::::       |||
Db     6279 ATCTCAAAAAGATTTGGAAATAACCAAAGAAAAT------CTGGCTCAAGCAGTTGAACA 6332
```

```
Qy   1397 rProGluCysValGlyAlaHis------------------LeuLeuProAlaGluLysSe 1411
             :::         |||                  ||||||         :::||
Db   6333 CCGCAAAAAGGCACAAGCAGAATTAGCTAGCTTCAAAGTCCTGCTAGATGACACTCAAAG 6392

Qy   1411 rSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyPr 1431
          |  |||                        :::       |||   :::   |||
Db   6393 TGAAGCAGCA------------------AGGGTCCTAGCAGACAATCTCAAGTTGAAAAA 6434

Qy   1431 oGluSerGlnAla------GluSerIleProIleIleValThrProAlaProGluSerTh 1449
             |||  |||:::      ||||||:::          :::          |||
Db   6435 GGAACTTCAGTCAAATAAAGAATCAGTTAAAAGCCAGATGAAACAAAAGGATGAA----- 6489

Qy   1449 rLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGl 1469
                     ||||||:::   :::    ::::::   :::       :::|||::
Db   6490 ---------GATCTTGAGCGAAGACTGGAACAGGCAGAAGAGAAGCACCTGAAAGAGAA 6539

Qy   1469 uAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysVa 1489
           :                        ||||||    :::|||        :::|||:::
Db   6540 GAAGAATATGCAAGAGAAACTGGATGCTTTGCGCAGAGAAAAAGTCCACTTGGAAGAGAC 6599

Qy   1489 lLeu-----------------------LysAlaGluProGluIleLeuGluLeuGluSe 1501
            :::                      |||   :::   |||:::   :::|||:::
Db   6600 AATTGGAGAGATTCAGGTTACTTTGAACAAGAAAGACAAGGAAGTTCAGCAACTTCAGGA 6659

Qy   1501 rLysSerAsnLysIleVal 1507
                   :::   |||
Db   6660 AAACTTGGACAGTACTGTG 6678
```

RESULT 14
US-09-724-676-7240
; Sequence 7240, Application US/09724676
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7240
;   LENGTH: 8189
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676-7240

```
Alignment Scores:
Pred. No.:               1.98e-09      Length:         8189
Score:                   415.00        Matches:        316
Percent Similarity:      38.17%        Conservative:   305
Best Local Similarity:   19.42%        Mismatches:     669
Query Match:             5.14%         Indels:         337
DB:                      5             Gaps:           60
```

US-09-902-432-4 (1-1596) x US-09-724-676-7240 (1-8189)

```
Qy     26 LeuSerGly-----------HisGlyProAlaAlaGluAlaSerGlyAlaAla------ 39
          |||||||||           ||||||         |||  ||||||   |||
Db   2404 CTCTCAGGTGAAGGAACTTAGCATGGTAACAGAATTGAGGGCTCAGGTAAAGCAACTGGA 2463

Qy     40 GlyAspPro-AlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSe 59
             :::|||  |||:::|||:::         :::|||     ::::::  |||    ::
Db   2464 AATGAACCTTGCAGAAGCAGAAAGGCAAAGAAGACTTGATTATGAAAGC---CAAACTGC 2520

Qy     59 rSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGl 79
           :      |||   :::   ||||||      :::|||                :::
Db   2521 CCATGACAACCTGCTCACTGAACAG------ATCCATAGTCTCAGCATAGAAGCCAAATC 2574

Qy     79 yGlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGl 99
           ::::::   ::::::          :::|||   |||   |||   :::||||||:::
Db   2575 TAAAGATGTGAAAATT---------GAAGTTTTACAGAATGAACTGGATGATGTGCAG-- 2623

Qy     99 uLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAs 119
                        :::   :::      ||||||   :::         :::
Db   2624 -------------CTTCAGTTTTCTGAGCAGAGTACCCTGATAAGAAGCCTGCAAAGCCA 2670

Qy    119 pGlyGlnGluGluThrSerGluIleIle---------GluGlnIleProAlaSerGluAs 136
             |||   :::   ||||||:::::::          |||::::::          ::
Db   2671 GCTGCAAAATAAGGAAAGTGAAGTGCTTGAGGGGGCAGAACGTGTAAGGCATATCTCAAG 2730

Qy    136 nAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLy 156
          :   |||||||||:::   |||         :::       ::::::
Db   2731 TAAAGTGGAAGAACTGTCCCAGGCTCTTTCACAGAAGGAACTTGAAATA----------- 2779

Qy    156 sValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAs 176
                                           |||  |||
Db   2780 -----------------------------ACAAAAATGGAT----------------- 2791

Qy    176 pThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGl 196
                ||||||||||      ||||||   :::   |||      :::   ::::::
Db   2792 -------CAGCTCTTACTAGAGAAAAAGAGAGATGTGGAAACCCTCCAACAAACCATCGA 2844

Qy    196 yAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGl 216
             |||   |||       :::      ::::::   |||           :::
Db   2845 GGAGAAGGATCAACAAGTGACAGAAATCAGCTTTAGTATGACTGAGAAAATGGTTCAGCT 2904

Qy    216 uSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSe 236
           :::|||   |||   |||      :::      |||||||||:::
Db   2905 TAATGAAGAGAAGTTTTCTCTTGGGGTTGAAATTAAGACTCTTAAAGAA----------- 2953

Qy    236 rThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGl 256
              ::::::   |||   :::   :::::::::|||   :::|||:::   ::::::::
Db   2954 ----CAGCTAAATTTATTATCCAGAGCTGAGGAAGCAAAAAAAGAGCAGGTGGAAGAAGA 3009

Qy    256 uGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProVa 276
           :   |||               :::               ||||||:::      ::
Db   3010 TAATGAAGTTTCTTCTGGCCTTAAACAAAATTATGATGAGATGAGCCCAGCAGGACAAAT 3069

Qy    276 lAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpAr 296
           ::::   |||                              |||   :::      ::
Db   3070 AAGTAAGGAAGAA--------------------CTTCAGCATGAATTTGACCTTCTGAA 3108

Qy    296 gLysLysThrSerPheLysLysSerLys----------------------GluAspAs 308
           :|||:::         :::|||   |||                       ::::::
Db   3109 GAAAGAAAATGAGCAGAGAAAGAGAAAGCTCCAGGCAGCTCTTATTAACAGAAAGGAGCT 3168

Qy    308 pLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLy 328
             |||:::      :::   :::|||:::   |||      ||||||   :::||
Db   3169 TCTGCAAAGAGTCAGTAGATTGGAAGAAGAATTAGCCAACTTGAAAGATGAATCTAAGAA 3228

Qy    328 sGluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaAr 348
           ||||            :::|||      |||      ||||||   :::
Db   3229 AGAAATCCCACTCAGTGAGACTGAGAGGGGAGAAGTGGAAGAAGATAAAGAAAACAAAGA 3288

Qy    348 gLeuSer---------------AlaAspTyrGluLysValGluLeuProLeuGluAspGl 363
              |||               :::      :::::::::|||:::   |||:::
Db   3289 ATACTCAGAAAAAATGTGTGACTTCTAAGTGCCAAGAAATAGAAATTTATTTAAAACAGAC 3348

Qy    363 nValGlyAspLeuGluAla--------------------SerSerGluGluLysCysAl 376
           :::   :::   |||                         |||||||||
Db   3349 AATATCTGAGAAAGAAGTGGAACTACAGCATATAAGGAAGGATTTGGAAGAAAAG----- 3403

Qy    376 aProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGl 396
              ||||||   |||                  |||      |||   :::|||   ::
Db   3404 ----CTGGCAGCTGAA-----------------GAGCAATTCCAGGCTCTGGTCAAACA 3441

Qy    396 uValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGl 416
          ::::::   :::      :::||||||   :::
Db   3442 GATGAATCAGACCTTGCAAGATAAAACAAACCAA------------------------- 3475

Qy    416 uGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluPr 436
                                        |||         :::::::::|||
Db   3476 ---------------------------ATAGATTTGCTCCAAGCAGAAATCAGTGAAAA 3507

Qy    436 oGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCy 456
            |||   :::   |||:::                     :::   :::
```

```
Db   3508 CCAAGCAATTATCCAGAAG--------------------TTAATCACAAGTAACACGGA 3546
Qy    456 sValSerGlyGlyAspHisThrGlnLeuThrAsp-----------LeuSerProGluGl 472
Db   3547 TGCAAGTGATGGGGACTCCGTAGCACTTGTAAAGGAAACAGTGGTGATAAGTCCACCTTG 3606
Qy    472 uLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGl 492
Db   3607 TACAGGTAGTAGTGAACACTGGAAACCAGAACTAGAAGAAAAGATACTGGCCCTTGAAAA 3666
Qy    492 nGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLy 512
Db   3667 AGAAAAGGAGCAACTTCAAAAGAAGCTACAGGAAGCCTTAACCTCCCGCAAGGCAATTCT 3726
Qy    512 sLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGl 532
Db   3727 TAAAAAGGCACAGGAGAAAGAAAGACATCTCAGG---------------GAGGAGCTAAA 3771
Qy    532 yGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSe 552
Db   3772 GCAACAGAAAGATGACTATAATCGCTTGCAAGAACAGTTTGATGAGCAAAGCAAGGAAAA 3831
Qy    552 r----------------------------------------------------SerAl 554
Db   3832 TGAGAATATTGGAGACCAGCTAAGGCAACTCCAGATTCAAGTAAGGGAATCCATAGACGG 3891
Qy    554 aSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaPr 574
Db   3892 AAAACTCCCAAGCACAGACCAGCAGGAATCGTGTTCTTCCACTCCAGGTTTAGAAGAACC 3951
Qy    574 oGlnAspGlyGluAlaGluGluGlyThrThr---------------------------- 584
Db   3952 TTTATTCAAAGCCACAGAACAGCATCACACTCAACCTGTTTTAGAGTCCAACTTGTGCCC 4011
Qy    585 ----------------SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSe 599
Db   4012 AGACTGGCCTTCTCATTCTGAAGATGCGAGTGCTCTGCAGGGCGGAACTTCTGTTGCCCA 4071
Qy    599 rPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGl 619
Db   4072 GATTAAGGCCCAGCTG--------------------------------------AAGGA 4092
Qy    619 uGluGluLeuGluLysValLys---SerAlaThrLeuSerSerThrAspSerThrValSe 638
Db   4093 AATAGAGGCTGAGAAAGTAGAGTTAGAATTGAAAGTTAGTTCTACAACAAGTGAGCTTAC 4152
Qy    638 rGluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysAr 658
Db   4153 TAAAAAATCAGAAGAGGTATTTCAGTTACAAGAGCAGATAAATAAACAGGGTTTAGAAAT 4212
Qy    658 gArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysAr 678
Db   4213 CGAGAGTCTAAAGACAGTATCCCATGAAGCTGAAGTCCATGCCGAAAGCCTGCAGCAGAA 4272
Qy    678 gAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSe 698
Db   4273 ATTG-----------------------GAAAGCAGCCAACTACAAATTGCTGGCCTAGA 4308
Qy    698 rHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSe 718
Db   4309 ACATCTAAGAGAATTGCAACCTAAACTGGATGAACTG---------CAAAAACTCATAAG 4359
Qy    718 rThrGlnGluGlnAspGlnAlaGlnGlySer---------SerSerProGluProAlaGl 735
Db   4360 CAAAAAGGAAGAAGACGTTAGCTACCTTTCTGGACAACTTAGTGAGAAAGAAGCAGCTCT 4419
Qy    735 ySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPr 755
Db   4420 CACTAAAATACAGACAGAGATAATAGAACAAGAAGATTTAATTAAGGCTCTGCATACA-- 4477
Qy    755 oArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLe 775
Db   4478 -------------CAGCTAGAAATGCAAGCCAAAGAGCATGATGAGAGGATAAAGCAGCT 4524
```

```
Qy    775 uSerThrGluIle--------------------GluProSerArgGluGluSerTrpVa 788
Db   4525 ACAGGTGGAACTTTGTGAAATGAAGCAAAAACCAGAAGAGATTGGAGAAGAAAGT----- 4579
Qy    788 lSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGl 808
Db   4580 ------------------------AGAGCAAAGCAACAAATACAAAGGAAACTGCAAGC 4614
Qy    808 nAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAl 828
Db   4615 TGCCCTTATTTCCCGAAAA---------GAAGCACTAAAAGAAAACAAAAGTCTCCAAGA 4665
Qy    828 aValValProLeuSerGluTyrAsnAlaValGluArg---------------------- 840
Db   4666 GGAATTGTCTTTGGCCAGA---GGTACCATTGAACGTCTCACCAAGTCTCTGGCAGATGT 4722
Qy    841 -GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVa 860
Db   4723 GGAAAGCCAAGTTTCTGCTCAAAATAAAGAAAAAGATACGGTCTTAGGAAGGTTAGCTCT 4782
Qy    860 lSerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyTh 880
Db   4783 TCTTCAAGAAGAAAGAGACAAACTCATTACAGAAATGGACAGGTCTTTATTGGAAAATCA 4842
Qy    880 rArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGl 900
Db   4843 GAGTCTCAGCAGCTCCTGTGAA------AGTCTAAAACTAGCTCTAGAGGGTCTTACTGA 4896
Qy    900 uProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAs 920
Db   4897 AGACAAGGAAAAGTTAGTGAAGGAAATT---------------GAATCTTTGAAATCTTC 4941
Qy    920 pIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHi 940
Db   4942 TAAGATTGCAGAAAGTACT------------------------GAGTGGCAAGAGAAACA 4977
Qy    940 sAspAsp--------------------MetValThrSerGluValAspPheThrSerGl 953
Db   4978 CAAGGAGCTACAAAAAGAGTATGAAATTCTTCTGCAGTCCTATGAGAATGTTAGTAATGA 5037
Qy    953 uAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGlu------------- 968
Db   5038 AGCAGAAAGGATTCAGCATGTGGTGGAAGCTGTGAGGCAAGAGAAACAAGAACTGTATGG 5097
Qy    969 ----------ValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaVa 985
Db   5098 CAAGTTAAGAAGCACAGAGGCAAACAAGAAGGAGACAGAAAAGCAGTTGCAGGAAGCTGA 5157
Qy    985 lSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGl 1005
Db   5158 GCAAGAAATGGAGGAAATGAAAGAAAAGATGAGAAAGTTT------GCTAAATCTAAACA 5211
Qy   1005 uSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaVa 1025
Db   5212 GCAGAAAATCCTAGAGCTGGAAGAAGAGAATGACCGGCTTAGGGCAGAGGTGCACCCTGC 5271
Qy   1025 lAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrGl 1045
Db   5272 AGGAGATACAGCTAAAGAG---------------TGTATGGAAACACTTCTTTCTTCCAA 5316
Qy   1045 ySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGl 1065
Db   5317 TGCCAGCATGAAGGAAGAACTTGAAAGGGTCAAAATGGAGTATGAAACCCTTTCTAAGAA 5376
Qy   1065 u---------------LysGluLysAspValMetProLysGlyProValGlnGluAlaGl 1080
Db   5377 GTTTCAGTCTTTAATGTCTGAGAAAGACTCTCTAAGTGAA---GAGGTTCAAGAT----- 5428
Qy   1080 yAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVa 1100
Db   5429 -TTAAAGCATCAGATAGAAGGTAATGTATCTAAACAAGCTAACCTAGAGGCCACCGAGAA 5487
```

```
Qy   1100 lProGluValThrAlaAspValAspHisValAlaThrCysGlnVal---------IleLy 1117
              :::      :::|||          |||      :::             ::
Db   5488 ACATGATAACCAAACGAATGTCACTGAAGAGGGAACACAGTCTATACCAGGTGAGACTGA 5547

Qy   1117 sLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSe 1137
           :   |||      :::     :::     |||      ||||||::::::        ::
Db   5548 AGAGCAAGACTCTCTGAGTATGAGCACAAGACCTACATGTTCAGAATCGGTTCCATCAGC 5607

Qy   1137 rGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAs 1157
          :::::::     :::   |||            |||   :::        :::    ::
Db   5608 GAAGAGT------GCCAACCCTGCTGTAAGTAAGGATTTCAGCTCACATGATGAAATTAA 5661

Qy   1157 p---------GluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSe 1174
          :         :::   ||||||             :::                :::
Db   5662 TAACTACCTACAGCAGATTGATCAGCTCAAAGAAAGAATTGCTGGATTAGAGGAGGAGAA 5721

Qy   1174 rGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAs 1194
             |||      :::|||   :::   |||   :::   |||:::   :::||||||   ::
Db   5722 GCAGAAAAACAAGGAATTTAGCCAGACTTTAGAAAATGAGAAA---AATACCTTACTGAG 5778

Qy   1194 nAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProTh 1214
           :   :::      ::::::      ||||||                |||:::
Db   5779 TCAGATATCAACAAAGGAT------GGTGAA------------------CTAAAAATGCT 5814

Qy   1214 rGlnGlnGluLeuThr--------------------------AlaAlaAlaValProVa 1225
           |||:::|||:::|||                                 :::
Db   5815 TCAGGAGGAAGTAACCAAAATGAACCTGTTAAATCAGCAAATCCAAGAAGAACTCTCCAG 5874

Qy   1225 lLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysVa 1245
            :::     |||    :::          |||    |||    |||   |||:::      ::
Db   5875 AGTTACCAAACTAAAGGAGACAGCAGAAGAAGAGAAAGATGATTTGGAAGAGAGGCTTAT 5934

Qy   1245 lLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspVa 1265
           :   :::         |||:::          :::|||              |||||
Db   5935 GAATCAATTAGCAGAACTT---------AATGGAAGCATTGGGAATTACTGTCAGGATGT 5985

Qy   1265 lThr-----------------------TyrAspSerGluValMetGlyValAlaGlyCy 1277
           ||||                        :::|||||||:::      :::          ||
Db   5986 TACAGATGCCCAAATAAAAAATGAGCTATTGGAATCTGAAATGAAGAACCTTAAAAAGTG 6045

Qy   1277 sGlnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluTh 1297
           |      :::          |||   |||   |||      |||:::   ::::::|||::
Db   6046 TGTGAGTGAATTGGAAGAAGAAAAGCAGCAGTTAGTCAAGGAAAAAACTAAGGTGGAATC 6105

Qy   1297 rAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGl 1317
          ::::::    ||||||      |||                        |||   |||    ::
Db   6106 AGAAATACGAAAGGAATATTTGGAGAAAATACAAGGTGCTCAGAAAGAACCCGGAAATAA 6165

Qy   1317 nGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMe 1337
          :                                  ::::::   |||   :::      |||
Db   6166 A--------------------------AGCCATGCAAAGGAACTTCAGGAACTGTTAAA 6198

Qy   1337 tProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAs 1357
                       :::      |||   |||                    :::
Db   6199 AGAAAAACAACAAGAAGTAAAGCAGCTA------------CAGAAGGACTGCATCAGGTA 6246

Qy   1357 pGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGl 1377
           |||:::|||          :::|||         |||::::::|||:::        |||::
Db   6247 TCAAGAGAAAATTAGTGCTCTGGAGAGAACTGTTAAAGCTCTAGAATTTGTTCAAACTGA 6306

Qy   1377 nThrAlaGluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSe 1397
          ::::   :::   :::|||         |||         :::::::::       |||
Db   6307 ATCTCAAAAAGATTTGGAAATAACCAAAGAAAAT------CTGGCTCAAGCAGTTGAACA 6360

Qy   1397 rProGluCysValGlyAlaHis------------------LeuLeuProAlaGluLysSe 1411
             :::          |||                         ||||||          :::||
Db   6361 CCGCAAAAAGGCACAAGCAGAATTAGCTAGCTTCAAAGTCCTGCTAGATGACACTCAAAG 6420

Qy   1411 rSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyPr 1431
           |   |||                         :::        |||   :::    |||
Db   6421 TGAAGCAGCA------------------AGGGTCCTAGCAGACAATCTCAAGTTGAAAAA 6462

Qy   1431 oGluSerGlnAla------GluSerIleProIleIleValThrProAlaProGluSerTh 1449
                 |||   |||:::       ||||||:::          :::           |||
Db   6463 GGAACTTCAGTCAAATAAAGAATCAGTTAAAAGCCAGATGAAACAAAAGGATGAA----- 6517

Qy   1449 rLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGl 1469
                    ||||||:::        :::       ::::::   :::      :::|||::
Db   6518 ----------GATCTTGAGCGAAGACTGGAACAGGCAGAAGAGAAGCACCTGAAAGAGAA 6567

Qy   1469 uAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysVa 1489
           :                      |||||         :::|||         :::|||:::
Db   6568 GAAGAATATGCAAGAGAAACTGGATGCTTTGCGCAGAGAAAAAGTCCACTTGGAAGAGAC 6627

Qy   1489 lLeu-----------------------LysAlaGluProGluIleLeuGluLeuGluSe 1501
            :::                            |||   :::   |||:::   :::|||:::
Db   6628 AATTGGAGAGATTCAGGTTACTTTGAACAAGAAAGACAAGGAAGTTCAGCAACTTCAGGA 6687

Qy   1501 rLysSerAsnLysIleVal 1507
                    :::      |||
Db   6688 AAACTTGGACAGTACTGTG 6706
```

RESULT 15
US-09-724-676A-7240
; Sequence 7240, Application US/09724676A
; GENERAL INFORMATION:
; APPLICANT: Compugen LTD
; TITLE OF INVENTION: Variants of alternative splicing
; FILE REFERENCE: 129181.4 Compugen
; CURRENT APPLICATION NUMBER: US/09/724,676A
; CURRENT FILING DATE: 2000-11-28
; NUMBER OF SEQ ID NOS: 97222
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 7240
;   LENGTH: 8189
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-724-676A-7240

Alignment Scores:

Pred. No.:	1.98e-09	Length:	8189
Score:	415.00	Matches:	316
Percent Similarity:	38.17%	Conservative:	305
Best Local Similarity:	19.42%	Mismatches:	669
Query Match:	5.14%	Indels:	337
DB:	5	Gaps:	60

US-09-902-432-4 (1-1596) x US-09-724-676A-7240 (1-8189)

```
Qy     26 LeuSerGly------------HisGlyProAlaAlaGluAlaSerGlyAlaAla------ 39
          |||||||||                ||||||         |||   ||||||   |||
Db   2404 CTCTCAGGTGAAGGAACTTAGCATGGTAACAGAATTGAGGGCTCAGGTAAAGCAACTGGA 2463

Qy     40 GlyAspPro-AlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSe 59
             :::|||  |||:::|||:::                :::|||      ::::::   |||   ::
Db   2464 AATGAACCTTGCAGAAGCAGAAAGGCAAAGAAGACTTGATTATGAAAGC---CAAACTGC 2520

Qy     59 rSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGl 79
           :       |||   :::   ||||||      :::|||                  :::
Db   2521 CCATGACAACCTGCTCACTGAACAG------ATCCATAGTCTCAGCATAGAAGCCAAATC 2574

Qy     79 yGlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGl 99
            ::::::    ::::::            :::|||   |||   |||    :::||||||:::
Db   2575 TAAAGATGTGAAAATT---------GAAGTTTTACAGAATGAACTGGATGATGTGCAG-- 2623

Qy     99 uLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAs 119
                      :::   :::         ||||||      :::         :::
Db   2624 -------------CTTCAGTTTTCTGAGCAGAGTACCCTGATAAGAAGCCTGCAAAGCCA 2670

Qy    119 pGlyGlnGluGluThrSerGluIleIle---------GluGlnIleProAlaSerGluAs 136
               |||   :::   ||||||::::::           |||::::::                ::
Db   2671 GCTGCAAAATAAGGAAAGTGAAGTGCTTGAGGGGGCAGAACGTGTAAGGCATATCTCAAG 2730

Qy    136 nAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLy 156
           :    |||||||||:::    |||                 :::       ::::::
Db   2731 TAAAGTGGAAGAACTGTCCCAGGCTCTTTCACAGAAGGAACTTGAAATA----------- 2779
```

```
Qy    156 sValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAs 176
                                            |||  |||
Db   2780 -------------------------------ACAAAAATGGAT----------------- 2791

Qy    176 pThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGl 196
                 ||||||||       ||||||  :::    |||      :::    ::::::
Db   2792 -------CAGCTCTTACTAGAGAAAAAGAGAGATGTGGAAACCCTCCAACAAACCATCGA 2844

Qy    196 yAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGl 216
                  |||   |||           :::     ::::::    |||        :::
Db   2845 GGAGAAGGATCAACAAGTGACAGAAATCAGCTTTAGTATGACTGAGAAAATGGTTCAGCT 2904

Qy    216 uSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSe 236
           :::|||   |||   |||         :::      ||||||||:::
Db   2905 TAATGAAGAGAAGTTTTCTCTTGGGGTTGAAATTAAGACTCTTAAAGAA----------- 2953

Qy    236 rThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGl 256
              ::::::   |||  :::   :::::::::|||  :::|||:::   ::::::::
Db   2954 ----CAGCTAAATTTATTATCCAGAGCTGAGGAAGCAAAAAAAGAGCAGGTGGAAGAAGA 3009

Qy    256 uGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProVa 276
          :   |||              :::                  ||||||:::     ::
Db   3010 TAATGAAGTTTCTTCTGGCCTTAAACAAAATTATGATGAGATGAGCCCAGCAGGACAAAT 3069

Qy    276 lAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpAr 296
          ::::   |||                              |||   :::         ::
Db   3070 AAGTAAGGAAGAA--------------------CTTCAGCATGAATTTGACCTTCTGAA 3108

Qy    296 gLysLysThrSerPheLysLysSerLys-----------------------GluAspAs 308
          :|||:::         :::|||   |||                        ::::::
Db   3109 GAAAGAAAATGAGCAGAGAAAGAGAAAGCTCCAGGCAGCTCTTATTAACAGAAAGGAGCT 3168

Qy    308 pLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLy 328
           |||:::         :::   :::|||:::   |||         |||||   :::||
Db   3169 TCTGCAAAGAGTCAGTAGATTGGAAGAAGAATTAGCCAACTTGAAAGATGAATCTAAGAA 3228

Qy    328 sGluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaAr 348
          ||||               :::|||        |||     ||||||   :::
Db   3229 AGAAATCCCACTCAGTGAGACTGAGAGGGGAGAAGTGGAAGAAGATAAAGAAAACAAAGA 3288

Qy    348 gLeuSer---------------AlaAspTyrGluLysValGluLeuProLeuGluAspGl 363
              |||              :::        :::::::::|||:::   |||:::
Db   3289 ATACTCAGAAAAATGTGTGACTTCTAAGTGCCAAGAAATAGAAATTTATTTAAAACAGAC 3348

Qy    363 nValGlyAspLeuGluAla--------------------SerSerGluGluLysCysAl 376
           :::   :::   |||                              ||||||||
Db   3349 AATATCTGAGAAAGAAGTGGAACTACAGCATATAAGGAAGGATTTGGAAGAAAAG----- 3403

Qy    376 aProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGl 396
              ||||||   |||               |||       |||   :::|||   ::
Db   3404 ----CTGGCAGCTGAA------------------GAGCAATTCCAGGCTCTGGTCAAACA 3441

Qy    396 uValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGl 416
          ::::::    :::      :::|||||   :::
Db   3442 GATGAATCAGACCTTGCAAGATAAAACAAACCAA------------------------- 3475

Qy    416 uGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluPr 436
                                            |||         :::::::::|||
Db   3476 ----------------------------ATAGATTTGCTCCAAGCAGAAATCAGTGAAAA 3507

Qy    436 oGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCy 456
           |||   :::   |||:::                      :::   :::
Db   3508 CCAAGCAATTATCCAGAAG---------------------TTAATCACAAGTAACACGGA 3546

Qy    456 sValSerGlyGlyAspHisThrGlnLeuThrAsp------------LeuSerProGluGl 472
              |||   ||||||          |||                   :::||||||
Db   3547 TGCAAGTGATGGGGACTCCGTAGCACTTGTAAAGGAAACAGTGGTGATAAGTCCACCTTG 3606

Qy    472 uLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGl 492
                        :::|||   :::      :::   |||      :::         ::
Db   3607 TACAGGTAGTAGTGAACACTGGAAACCAGAACTAGAAGAAAAGATACTGGCCCTTGAAAA 3666

Qy    492 nGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLy 512
          :|||:::   ::::::|||            :::   |||   |||         :::
Db   3667 AGAAAAGGAGCAACTTCAAAAGAAGCTACAGGAAGCCTTAACCTCCCGCAAGGCAATTCT 3726

Qy    512 sLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGl 532
           |||   :::   :::|||::::::      |||              ||||||
Db   3727 TAAAAAGGCACAGGAGAAAGAAAGACATCTCAGG---------------GAGGAGCTAAA 3771

Qy    532 yGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSe 552
           :::   :::      :::            |||      ||||||||      |||::
Db   3772 GCAACAGAAAGATGACTATAATCGCTTGCAAGAACAGTTTGATGAGCAAAGCAAGGAAAA 3831

Qy    552 r------------------------------------------------------SerAl 554
          :
Db   3832 TGAGAATATTGGAGACCAGCTAAGGCAACTCCAGATTCAAGTAAGGGAATCCATAGACGG 3891

Qy    554 aSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaPr 574
                 |||         ::::::   :::|||               ||||||   ||
Db   3892 AAAACTCCCAAGCACAGACCAGCAGGAATCGTGTTCTTCCACTCCAGGTTTAGAAGAACC 3951

Qy    574 oGlnAspGlyGluAlaGluGluGlyThrThr----------------------------- 584
          |               |||:::     |||
Db   3952 TTTATTCAAAGCCACAGAACAGCATCACACTCAACCTGTTTTAGAGTCCAACTTGTGCCC 4011

Qy    585 ----------------SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSe 599
                          |||:::            :::|||   |||      |||
Db   4012 AGACTGGCCTTCTCATTCTGAAGATGCGAGTGCTCTGCAGGGCGGAACTTCTGTTGCCCA 4071

Qy    599 rPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGl 619
               |||   :::                                       |||||
Db   4072 GATTAAGGCCCAGCTG---------------------------------------AAGGA 4092

Qy    619 uGluGluLeuGluLysValLys---SerAlaThrLeuSerSerThrAspSerThrValSe 638
          |   |||   |||||||||:::            :::||||||||   |||   :::::
Db   4093 AATAGAGGCTGAGAAAGTAGAGTTAGAATTGAAAGTTAGTTCTACAACAAGTGAGCTTAC 4152

Qy    638 rGluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysAr 658
          ::::      :::||||||      :::   |||:::          :::   :::
Db   4153 TAAAAAATCAGAAGAGGTATTTCAGTTACAAGAGCAGATAAATAAACAGGGTTTAGAAAT 4212

Qy    658 gArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysAr 678
                       :::||||||   ||||||   :::          |||   ::::::::
Db   4213 CGAGAGTCTAAAGACAGTATCCCATGAAGCTGAAGTCCATGCCGAAAGCCTGCAGCAGAA 4272

Qy    678 gAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSe 698
          :                           |||                :::   |||
Db   4273 ATTG-----------------------GAAAGCAGCCAACTACAAATTGCTGGCCTAGA 4308

Qy    698 rHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSe 718
           |||      |||         |||      |||                        ||
Db   4309 ACATCTAAGAGAATTGCAACCTAAACTGGATGAACTG---------CAAAAACTCATAAG 4359

Qy    718 rThrGlnGluGlnAspGlnAlaGlnGlySer---------SerSerProGluProAlaGl 735
          |   :::|||:::|||   :::      |||         |||      |||   |||
Db   4360 CAAAAAGGAAGAAGACGTTAGCTACCTTTCTGGACAACTTAGTGAGAAAGAAGCAGCTCT 4419

Qy    735 ySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPr 755
           :::      :::   |||   :::            :::      |||   |||   |||
Db   4420 CACTAAAATACAGACAGAGATAATAGAACAAGAAGATTTAATTAAGGCTCTGCATACA-- 4477

Qy    755 oArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLe 775
                       :::   :::::::::   |||      |||      :::::||||
Db   4478 -------------CAGCTAGAAATGCAAGCCAAAGAGCATGATGAGAGGATAAAGCAGCT 4524

Qy    775 uSerThrGluIle---------------------GluProSerArgGluGluSerTrpVa 788
          |      |||:::                     |||         ||||||||
Db   4525 ACAGGTGGAACTTTGTGAAATGAAGCAAAAACCAGAAGAGATTGGAGAAGAAAGT----- 4579

Qy    788 lSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGl 808
                                   |||   |||::::::         |||   :::
Db   4580 ------------------------AGAGCAAAGCAACAAATACAAAGGAAACTGCAAGC 4614

Qy    808 nAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAl 828
```

```
Db   4615 TGCCCTTATTTCCCGAAAA---------GAAGCACTAAAAGAAAACAAAAGTCTCCAAGA 4665

Qy    828 aValValProLeuSerGluTyrAsnAlaValGluArg---------------------- 840

Db   4666 GGAATTGTCTTTGGCCAGA---GGTACCATTGAACGTCTCACCAAGTCTCTGGCAGATGT 4722

Qy    841 -GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVa 860

Db   4723 GGAAAGCCAAGTTTCTGCTCAAAATAAAGAAAAAGATACGGTCTTAGGAAGGTTAGCTCT 4782

Qy    860 lSerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyTh 880

Db   4783 TCTTCAAGAAGAAAGAGACAAACTCATTACAGAAATGGACAGGTCTTTATTGGAAAATCA 4842

Qy    880 rArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGl 900

Db   4843 GAGTCTCAGCAGCTCCTGTGAA------AGTCTAAAACTAGCTCTAGAGGGTCTTACTGA 4896

Qy    900 uProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAs 920

Db   4897 AGACAAGGAAAAGTTAGTGAAGGAAATT---------------GAATCTTTGAAATCTTC 4941

Qy    920 pIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHi 940

Db   4942 TAAGATTGCAGAAAGTACT-----------------------GAGTGGCAAGAGAAACA 4977

Qy    940 sAspAsp--------------------MetValThrSerGluValAspPheThrSerGl 953

Db   4978 CAAGGAGCTACAAAAAGAGTATGAAATTCTTCTGCAGTCCTATGAGAATGTTAGTAATGA 5037

Qy    953 uAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGlu------------- 968

Db   5038 AGCAGAAAGGATTCAGCATGTGGTGGAAGCTGTGAGGCAAGAGAAACAAGAACTGTATGG 5097

Qy    969 ---------ValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaVa 985

Db   5098 CAAGTTAAGAAGCACAGAGGCAAACAAGAAGGAGACAGAAAAGCAGTTGCAGGAAGCTGA 5157

Qy    985 lSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGl 1005

Db   5158 GCAAGAAATGGAGGAAATGAAAGAAAAGATGAGAAAGTTT------GCTAAATCTAAACA 5211

Qy   1005 uSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaVa 1025

Db   5212 GCAGAAAATCCTAGAGCTGGAAGAAGAGAATGACCGGCTTAGGGCAGAGGTGCACCCTGC 5271

Qy   1025 lAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrGl 1045

Db   5272 AGGAGATACAGCTAAAGAG---------------TGTATGGAAACACTTCTTTCTTCCAA 5316

Qy   1045 ySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGl 1065

Db   5317 TGCCAGCATGAAGGAAGAACTTGAAAGGGTCAAAATGGAGTATGAAACCCTTTCTAAGAA 5376

Qy   1065 u--------------LysGluLysAspValMetProLysGlyProValGlnGluAlaGl 1080

Db   5377 GTTTCAGTCTTTAATGTCTGAGAAAGACTCTCTAAGTGAA---GAGGTTCAAGAT----- 5428

Qy   1080 yAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVa 1100

Db   5429 -TTAAAGCATCAGATAGAAGGTAATGTATCTAAACAAGCTAACCTAGAGGCCACCGAGAA 5487

Qy   1100 lProGluValThrAlaAspValAspHisValAlaThrCysGlnVal---------IleLy 1117

Db   5488 ACATGATAACCAAACGAATGTCACTGAAGAGGGAACACAGTCTATACCAGGTGAGACTGA 5547

Qy   1117 sLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSe 1137

Db   5548 AGAGCAAGACTCTCTGAGTATGAGCACAAGACCTACATGTTCAGAATCGGTTCCATCAGC 5607

Qy   1137 rGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAs 1157

Db   5608 GAAGAGT------GCCAACCCTGCTGTAAGTAAGGATTTCAGCTCACATGATGAAATTAA 5661

Qy   1157 p---------GluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSe 1174

Db   5662 TAACTACCTACAGCAGATTGATCAGCTCAAAGAAAGAATTGCTGGATTAGAGGAGGAGAA 5721

Qy   1174 rGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAs 1194

Db   5722 GCAGAAAAACAAGGAATTTAGCCAGACTTTAGAAAATGAGAAA---AATACCTTACTGAG 5778

Qy   1194 nAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProTh 1214

Db   5779 TCAGATATCAACAAAGGAT------GGTGAA------------------CTAAAAATGCT 5814

Qy   1214 rGlnGlnGluLeuThr--------------------------AlaAlaAlaValProVa 1225

Db   5815 TCAGGAGGAAGTAACCAAAATGAACCTGTTAAATCAGCAAATCCAAGAAGAACTCTCCAG 5874

Qy   1225 lLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysVa 1245

Db   5875 AGTTACCAAACTAAAGGAGACAGCAGAAGAAGAGAAAGATGATTTGGAAGAGAGGCTTAT 5934

Qy   1245 lLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspVa 1265

Db   5935 GAATCAATTAGCAGAACTT---------AATGGAAGCATTGGGAATTACTGTCAGGATGT 5985

Qy   1265 lThr-----------------------TyrAspSerGluValMetGlyValAlaGlyCy 1277

Db   5986 TACAGATGCCCAAATAAAAAATGAGCTATTGGAATCTGAAATGAAGAACCTTAAAAAGTG 6045

Qy   1277 sGlnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluTh 1297

Db   6046 TGTGAGTGAATTGGAAGAAGAAAAGCAGCAGTTAGTCAAGGAAAAAACTAAGGTGGAATC 6105

Qy   1297 rAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGl 1317

Db   6106 AGAAATACGAAAGGAATATTTGGAGAAAATACAAGGTGCTCAGAAAGAACCCGGAAATAA 6165

Qy   1317 nGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMe 1337

Db   6166 A--------------------------AGCCATGCAAAGGAACTTCAGGAACTGTTAAA 6198

Qy   1337 tProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAs 1357

Db   6199 AGAAAAACAACAAGAAGTAAAGCAGCTA------------CAGAAGGACTGCATCAGGTA 6246

Qy   1357 pGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGl 1377

Db   6247 TCAAGAGAAAATTAGTGCTCTGGAGAGAACTGTTAAAGCTCTAGAATTTGTTCAAACTGA 6306

Qy   1377 nThrAlaGluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSe 1397

Db   6307 ATCTCAAAAAGATTTGGAAATAACCAAAGAAAAT------CTGGCTCAAGCAGTTGAACA 6360

Qy   1397 rProGluCysValGlyAlaHis------------------LeuLeuProAlaGluLysSe 1411

Db   6361 CCGCAAAAAGGCACAAGCAGAATTAGCTAGCTTCAAAGTCCTGCTAGATGACACTCAAAG 6420

Qy   1411 rSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyPr 1431

Db   6421 TGAAGCAGCA------------------AGGGTCCTAGCAGACAATCTCAAGTTGAAAAA 6462

Qy   1431 oGluSerGlnAla------GluSerIleProIleIleValThrProAlaProGluSerTh 1449

Db   6463 GGAACTTCAGTCAAATAAAGAATCAGTTAAAAGCCAGATGAAACAAAAGGATGAA----- 6517

Qy   1449 rLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGl 1469

Db   6518 ----------GATCTTGAGCGAAGACTGGAACAGGCAGAAGAGAAGCACCTGAAAGAGAA 6567

Qy   1469 uAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysVa 1489

Db   6568 GAAGAATATGCAAGAGAAACTGGATGCTTTGCGCAGAGAAAAAGTCCACTTGGAAGAGAC 6627
```

```
Qy   1489 lLeu------------------------LysAlaGluProGluIleLeuGluLeuGluSe 1501
           :::                         |||   :::   |||:::   :::|||:::
Db   6628 AATTGGAGAGATTCAGGTTACTTTGAACAAGAAAGACAAGGAAGTTCAGCAACTTCAGGA 6687

Qy   1501 rLysSerAsnLysIleVal 1507
                 :::      |||
Db   6688 AAACTTGGACAGTACTGTG 6706
```

Search completed: December 13, 2002, 06:34:53
Job time : 2025 secs

```
OM nucleic - nucleic search, using sw model

Run on:         December 12, 2002, 14:13:30 ; Search time 14973 Seconds
                                              (without alignments)
                                              11973.104 Million cell updates/sec

Title:          US-09-902-432-3
Perfect score:  6160
Sequence:       1 cgggccggcgggagtagaag..........aaataaacatatttgactac 6160

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters:      4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       GenEmbl:*
                1:  gb_ba:*
                2:  gb_htg:*
                3:  gb_in:*
                4:  gb_om:*
                5:  gb_ov:*
                6:  gb_pat:*
                7:  gb_ph:*
                8:  gb_pl:*
                9:  gb_pr:*
                10:  gb_ro:*
                11:  gb_sts:*
                12:  gb_sy:*
                13:  gb_un:*
                14:  gb_vi:*
                15:  em_ba:*
                16:  em_fun:*
                17:  em_hum:*
                18:  em_in:*
                19:  em_mu:*
                20:  em_om:*
                21:  em_or:*
                22:  em_ov:*
                23:  em_pat:*
                24:  em_ph:*
                25:  em_pl:*
                26:  em_ro:*
                27:  em_sts:*
                28:  em_un:*
                29:  em_vi:*
                30:  em_htg_hum:*
                31:  em_htg_inv:*
                32:  em_htg_other:*
                33:  em_htg_mus:*
                34:  em_htg_pln:*
                35:  em_htg_rod:*
                36:  em_htg_mam:*
                37:  em_htg_vrt:*
                38:  em_sy:*
                39:  em_htgo_hum:*
                40:  em_htgo_mus:*
                41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6160	100.0	6160	10	RNU23146	U23146 Rattus norv
	2	5204.2	84.5	5236	10	RNU41453	U41453 Rattus norv
	3	4831.4	78.4	5134	6	AR071282	AR071282 Sequence
	4	4559.6	74.0	6197	10	AF326228	AF326228 Mus muscu
	5	4558	74.0	6195	10	AB020886	AB020886 Mus muscu
	6	4441.8	72.1	178954	2	AC115647	AC115647 Rattus no
	7	4304	69.9	5902	10	AF326230	AF326230 Mus muscu
	8	2179.6	35.4	6614	6	AX405857	AX405857 Sequence
	9	2144.4	34.8	6608	6	AX329766	AX329766 Sequence
	10	2144.4	34.8	6608	6	AX410793	AX410793 Sequence
	11	2144.4	34.8	6608	9	HSU81607	U81607 Homo sapien
	12	2136.8	34.7	6297	9	AF001504	AF001504 Homo sapi
	13	2118.8	34.4	6605	6	AR002603	AR002603 Sequence
	14	2089.6	33.9	6287	9	AB003476	AB003476 Homo sapi
	15	1906.4	30.9	104939	9	AL590413	AL590413 Human DNA
	16	1906.4	30.9	171158	2	AC023201	AC023201 Homo sapi
c	17	923.8	15.0	102822	2	AC097731	AC097731 Rattus no
c	18	911.8	14.8	178954	2	AC115647	AC115647 Rattus no
	19	681.2	11.1	129179	2	AC117205	AC117205 Mus muscu
	20	651	10.6	102822	2	AC097731	AC097731 Rattus no
c	21	550	8.9	563	6	AX401547	AX401547 Sequence
	22	466.4	7.6	1027	9	BC000188	BC000188 Homo sapi
	23	466.4	7.6	1027	9	BC022814	BC022814 Homo sapi
	24	416	6.8	8439	5	AF308810	AF308810 Xenopus l
	25	313.2	5.1	1981	9	HUMGRAVIN	M96322 Human gravi
	26	232.8	3.8	1058	11	G06673	G06673 human STS W
c	27	163.4	2.7	129179	2	AC117205	AC117205 Mus muscu
	28	162.2	2.6	178	6	AX401799	AX401799 Sequence
	29	162.2	2.6	178	10	RNU75404	U75404 Rattus norv
	30	160.6	2.6	696	9	HUMZD39H03	AF086250 Homo sapi
c	31	131.8	2.1	386	11	G21790	G21790 human STS W
c	32	108.6	1.8	179145	9	AC009785	AC009785 Homo sapi
	33	107.8	1.7	195690	10	AC098726	AC098726 Mus muscu
	34	104.8	1.7	177681	2	AC116238	AC116238 Rattus no
c	35	104.6	1.7	174990	10	AC027653	AC027653 Mus muscu
	36	101.6	1.6	35848	2	AC011553	AC011553 Homo sapi
	37	101.2	1.6	33082	2	AC091365	AC091365 Rattus no
c	38	101	1.6	193363	2	AL627104	AL627104 Mus muscu
c	39	99.8	1.6	7218	6	I66494	I66494 Sequence 14
c	40	99.8	1.6	68100	2	AC105787	AC105787 Rattus no
	41	99.6	1.6	196194	2	AC104920	AC104920 Mus muscu
c	42	99.2	1.6	206497	10	AC084389	AC084389 Mus muscu
c	43	98.8	1.6	45632	2	AC098158	AC098158 Rattus no
	44	98.6	1.6	125816	10	AL591127	AL591127 Mouse DNA
	45	98.2	1.6	194334	2	AC101527	AC101527 Mus muscu

ALIGNMENTS

```
RESULT 1
RNU23146
LOCUS       RNU23146                6160 bp    DNA     linear   ROD 19-OCT-1995
DEFINITION  Rattus norvegicus mitogenic regulation SSeCKS (322) gene, complete
            cds.
ACCESSION   U23146
VERSION     U23146.1  GI:1022682
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 6160)
  AUTHORS   Lin,X., Nelson,P.J., Frankfort,B., Tombler,E., Johnson,R. and
```

```
            Gelman,I.H.
  TITLE     Isolation and characterization of a novel mitogenic regulatory
            gene, 322, which is transcriptionally suppressed in cells
            transformed by src and ras
  JOURNAL   Mol. Cell. Biol. 15 (5), 2754-2762 (1995)
  MEDLINE   95257957
   PUBMED   7739556
REFERENCE   2  (bases 1 to 6160)
  AUTHORS   Gelman,I.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-1995) Irwin H. Gelman, Microbiology, Mount Sinai
            School of Medicine, One Gustave L. Levy Place, New York, NY 10029,
            USA
COMMENT     On Oct 19, 1995 this sequence version replaced gi:733449.
FEATURES             Location/Qualifiers
     source          1..6160
                     /organism="Rattus norvegicus"
                     /strain="Fisher"
                     /db_xref="taxon:10116"
                     /clone="13.2.2"
                     /cell_line="Rat-6"
                     /cell_type="fibroblast"
                     /tissue_type="embryo"
                     /clone_lib="R1-a (Robert Kraus, Mount Sinai School of
                     Medicine)"
     5'UTR           1..32
     gene            33..5044
                     /gene="322"
     CDS             33..4823
                     /gene="322"
                     /function="possible mitogenic regulatory gene"
                     /note="Src Suppressed C Kinase Substrate"
                     /citation=[1]
                     /codon_start=1
                     /evidence=experimental
                     /product="SSeCKS"
                     /protein_id="AAA79517.1"
                     /db_xref="GI:1022683"
                     /translation="MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPAD
                     ADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR
                     VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKF
                     VGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESE
                     LKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPV
                     NSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPA
                     SEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQE
                     VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELM
                     KSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSS
                     SGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL
                     EKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEK
                     VKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA
                     SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEG
                     EGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRR
                     KKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLG
                     AVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVE
                     EVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEA
                     SGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK
                     EESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGS
                     ETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLA
                     DSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGE
                     EPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKA
                     ADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA
                     APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQT
                     AEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAE
                     SIPIIVTPAPESTLHPDLQGEISASQRERSEEDKPDAGPDADGKESTAIEKVLKAEP
                     EILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTK
                     MKDAKMKHPVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK"
     misc_signal     1581..1590
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     misc_signal     1797..1811
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     misc_signal     2061..2073
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     misc_signal     2301..2313
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     misc_binding    4230..4358
                     /gene="322"
                     /note="encodes zinc-finger domain"
                     /citation=[1]
                     /bound_moiety="DNA, RNA or protein binding domain"
     3'UTR           4816..6160
     misc_signal     5032..5044
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     polyA_signal    5760..5765
     polyA_signal    6136..6141
     polyA_site      6160
BASE COUNT     1853 a   1409 c   1753 g   1145 t
ORIGIN

  Query Match             100.0%;  Score 6160;  DB 10;  Length 6160;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 6160;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy        361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600

Qy        601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
```

```
Qy    661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy    721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy    781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy    841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy    901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy    961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy   1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy   1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy   1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy   1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy   1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy   1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy   1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy   1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy   1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy   1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy   1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy   1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy   1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy   1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy   1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy   1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy   1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy   2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy   2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy   2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy   2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy   2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy   2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400

Qy   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460

Qy   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520

Qy   2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580

Qy   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760

Qy   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120

Qy   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180

Qy   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380

Qy   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy   4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy   4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy   4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy   4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860

Qy   4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920

Qy   4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980

Qy   4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
```

```
Qy   5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100

Qy   5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160

Qy   5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220

Qy   5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280

Qy   5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340

Qy   5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400

Qy   5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460

Qy   5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520

Qy   5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAA 5580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAA 5580

Qy   5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640

Qy   5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700

Qy   5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760

Qy   5761 CAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTC 5820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5761 CAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTC 5820

Qy   5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880

Qy   5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940

Qy   5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000

Qy   6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTC 6060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTC 6060

Qy   6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120

Qy   6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
          ||||||||||||||||||||||||||||||||||||||||
Db   6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
```

```
RESULT 2
RNU41453
LOCUS       RNU41453                5236 bp    mRNA    linear   ROD 12-JAN-1999
DEFINITION  Rattus norvegicus PKC binding protein and substrate mRNA, complete
            cds.
ACCESSION   U41453
VERSION     U41453.1  GI:4096819
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 5236)
  AUTHORS   Chapline,C., Mousseau,B., Ramsay,K. and Jaken,S.
  TITLE     Identification of a major protein kinase C binding protein and
            substrate in rat embryo fibroblasts: Decreased expression in
            transformed cells
  JOURNAL   J. Biol. Chem. (1995) In press
REFERENCE   2  (bases 1 to 5236)
  AUTHORS   Chapline,C., Mousseau,B., Ramsay,K. and Jaken,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-1995) Susan Jaken, W Alton Jones Cell Science
            Center, Old Barn Road, Lake Placid, NY 12946, USA
FEATURES             Location/Qualifiers
     source          1..5236
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="72"
     CDS             33..4823
                     /note="major PKC binding protein in REF52 cells"
                     /codon_start=1
                     /product="PKC binding protein and substrate"
                     /protein_id="AAD03788.1"
                     /db_xref="GI:4096820"
                     /translation="MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPAD
                     ADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR
                     VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKF
                     VGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESE
                     LKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPV
                     NSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPA
                     SEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQE
                     VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELM
                     KSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSS
                     SGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL
                     EKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEK
                     VKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA
                     SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEG
                     EGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRR
                     KKRADGKQEQATVEDSGPVEINEDDPDVPAVVPLSEYDAVEREKMEAQGNAELPQLLG
                     AVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVE
                     EVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEA
                     SGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEERQTQAILQAVADKVK
                     EESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGS
                     ETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLA
                     DSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGE
                     EPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKA
                     ADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA
                     APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQT
                     AEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAE
                     SIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEP
                     EILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPAMQADSREPNRCWTK
                     MKVAKMKHPVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK"
BASE COUNT     1622 a   1206 c   1583 g    825 t
ORIGIN

  Query Match             84.5%;  Score 5204.2;  DB 10;  Length 5236;
  Best Local Similarity   99.7%;  Pred. No. 0;
  Matches 5223;  Conservative    0;  Mismatches   13;  Indels    1;  Gaps    1;
```

```
Qy        1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy       61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy      121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy      181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy      241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy      301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy      361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy      421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy      481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy      541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600

Qy      601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660

Qy      661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy      721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy      781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy      841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy      901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy      961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy     1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy     1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy     1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy     1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy     1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy     1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy     1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy     1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy     1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy     1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy     1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy     1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy     1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy     1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy     1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy     1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy     1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy     2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy     2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy     2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy       2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy       2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy       2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400

Qy       2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
              |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db       2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAGAACAAGCCA 2460

Qy       2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
              |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db       2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTGATGTCCCAGCCGTCG 2520

Qy       2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
              ||||||||||||||||| |||||||||||||||||||||||||||||||||||||| |||
Db       2521 TGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATGCGG 2580

Qy       2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy       2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy       2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760

Qy       2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy       2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy       2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy       2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy       3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy       3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120

Qy       3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180

Qy       3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy       3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy       3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy       3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy       3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy       3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy       3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy       3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy       3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy       3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy       3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy       3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy       3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy       3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy       4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy       4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy       4141 GCCTGGACACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4141 GCCTGGACACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy       4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy       4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy       4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
```

101

```
Qy    4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy    4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy    4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy    4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy    4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
           ||||||||||||||||||||||||||     |  ||||||||||||||||||||||||||
Db    4621 ATGATTCACAGACCCAGGTTCCTGCAATGCAGGCTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy    4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db    4681 GGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy    4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy    4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGAT-CAAA 4859

Qy    4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4860 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4919

Qy    4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4920 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4979

Qy    4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4980 ATCGCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5039

Qy    5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5040 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5099

Qy    5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5100 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5159

Qy    5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5160 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5219

Qy    5221 TCCAATGATTTCCACCC 5237
           |||||||||||||||||
Db    5220 TCCAATGATTTCCACCC 5236


RESULT 3
AR071282
LOCUS       AR071282                 5134 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 1 from patent US 5910442.
ACCESSION   AR071282
VERSION     AR071282.1  GI:7222170
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5134)
  AUTHORS   Gelman,I.H.
  TITLE     Tumor suppressor gene
  JOURNAL   Patent: US 5910442-A 1 08-JUN-1999;
FEATURES             Location/Qualifiers
     source          1..5134
                     /organism="unknown"
BASE COUNT     1511 a   1167 c   1451 g   1005 t
ORIGIN

  Query Match             78.4%;  Score 4831.4;  DB 6;  Length 5134;
  Best Local Similarity   97.9%;  Pred. No. 0;
  Matches 5048;  Conservative    0;  Mismatches   76;  Indels   31;  Gaps   14;

Qy    1016 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 1075
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 60

Qy    1076 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 1135
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 120

Qy    1136 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 1195
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 180

Qy    1196 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 1255
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 240

Qy    1256 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 1315
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 300

Qy    1316 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 1375
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 360

Qy    1376 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 1435
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 420

Qy    1436 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 1495
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 480

Qy    1496 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 1555
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 540

Qy    1556 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 1615
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 600

Qy    1616 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 1675
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 660

Qy    1676 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 1735
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 720

Qy    1736 AGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGA 1795
           ||||||||||||||| ||||||||||||||||||||||||||||||||||.   |  ||
Db     721 AGGCCGCTGGAAGCA-CCCAGGATGGGGAAGCTGAGGAAGGAACTACTTC----GTGGA 775

Qy    1796 GAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 1855
           ||  |  |||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db     776 GAGAAGAAGAGGAAGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 835

Qy    1856 ACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 1915
           || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     836 AC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 894
```

```
Qy   1916 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 1975
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    895 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 954

Qy   1976 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 2035
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    955 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 1014

Qy   2036 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGA 2095
          ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db   1015 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCA---GATAT 1071

Qy   2096 AGGAGGGCCAAGGACACTGGGA-GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 2154
          | |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   1072 AAGAGGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 1131

Qy   2155 AAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 2214
          |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1132 AAGAAGCC-GAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 1190

Qy   2215 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 2274
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1191 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 1250

Qy   2275 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCG 2334
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db   1251 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGAAG 1310

Qy   2335 AAGACTCTAGTGTAG--------AGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGA 2386
            |     | | |||            |||||||||||||||||||||| ||||||||||
Db   1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy   2387 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA 2446
          |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy   2447 -GCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTA 2505
           |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db   1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy   2506 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAG 2565
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy   2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
          |||||||||| ||||||||||           |||    |||    |||||||||||||||
Db   1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGT---AGTGTCCGAGGAGCTCA 1607

Qy   2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy   2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy   2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy   2806 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 2865
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy   2866 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 2925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy   2926 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 2985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy   2986 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 3045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   3046 AGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 3105
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   3106 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 3165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   3166 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 3225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   3226 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 3285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   3286 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGG 3345
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   3346 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 3405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   3406 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 3465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   3466 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 3525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   3526 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 3585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   3586 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 3645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   3646 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCG 3705
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   3706 TTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 3765
          |   |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   3766 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 3825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   3826 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 3885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   3886 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 3945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   3946 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 4005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   4006 AAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 4065
          || | | |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   4066 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 4125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy    4126 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 4185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy    4186 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 4245
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy    4246 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 4305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy    4306 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 4365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy    4366 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 4425
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy    4426 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 4485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy    4486 CAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 4545
           ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy    4546 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 4605
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy    4606 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGG 4665
           |||||||||||||||||||||||||||||||||||||||||    |  | ||||||||||
Db    3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy    4666 AGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
           |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db    3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy    4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCT 4785
           |||||||||||||||||||||||||||||||||||||||||       ||||||||||||
Db    3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGCTCAG--CTCGGAAATGCT 3765

Qy    4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 3825

Qy    4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
           |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db    3826 ACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 3884

Qy    4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
           |||||||    |  ||||||||||||||||||||||||||||||||||||||||||||||
Db    3885 CAGAGGCA---GTGTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 3941

Qy    4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
           ||||||||||| |    |||||||||||||||||||||||||||||||||||||||||||
Db    3942 TAACCGAGGAGCG-ATGCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 4000

Qy    5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 5085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4001 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 4060

Qy    5086 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 5145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4061 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 4120

Qy    5146 TCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATC 5205
           |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    4121 TCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCATGAGATC 4179

Qy    5206 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 5265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4180 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 4239

Qy    5266 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4240 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4299

Qy    5326 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5385
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4300 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4359

Qy    5386 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 5445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4360 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 4419

Qy    5446 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 5505
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4420 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 4479

Qy    5506 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 5565
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4480 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 4539

Qy    5566 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 5625
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4540 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 4599

Qy    5626 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 5685
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4600 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 4659

Qy    5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 5745
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4660 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 4719

Qy    5746 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 5805
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4720 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 4779

Qy    5806 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 5865
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4780 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 4839

Qy    5866 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 5925
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4840 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 4899

Qy    5926 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 5985
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4900 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 4959

Qy    5986 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 6045
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4960 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 5019

Qy    6046 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 6105
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5020 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 5079

Qy    6106 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5080 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 5134
```

```
RESULT 4
AF326228
LOCUS       AF326228                6197 bp    mRNA    linear   ROD 16-JUL-2001
DEFINITION  Mus musculus PKC binding protein SSeCKS mRNA, complete cds.
ACCESSION   AF326228
VERSION     AF326228.1  GI:13236403
KEYWORDS    .
```

```
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 6197)
  AUTHORS   Camus,A., Mesbah,K., Rallu,M., Babinet,C. and Barra,J.
  TITLE     Gene trap insertion reveals two open reading frames in the mouse
            SSeCKS gene: the form predominantly detected in the nervous system
            is suppressed by the insertion while the other, specific of the
            testis, remains expressed
  JOURNAL   Mech. Dev. 105 (1-2), 79-91 (2001)
  MEDLINE   21322683
   PUBMED   11429284
REFERENCE   2  (bases 1 to 6197)
  AUTHORS   Camus,A., Mesbah,K., Babinet,C. and Barra,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2000) Immunologie, Institut Pasteur, 25 rue du
            Docteur Roux, Paris 75015, France
FEATURES             Location/Qualifiers
     source          1..6197
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="10"
     CDS             140..5194
                     /codon_start=1
                     /product="PKC binding protein SSeCKS"
                     /protein_id="AAK16150.1"
                     /db_xref="GI:13236404"
                     /translation="MGAGSSTEQRSPEQPAESDTPSELELSGHGPAAEASGAAGDPAD
                     ADPATKLPQKNGQLSAVNGVAEQEDVHVQEESQDGQEEEVTVEDVGQRESEDVKEKDR
                     AKEMAASSTVVEDITKDEQEETPEIIEQIPASESNVEEMAQAAESQANDVGFKKVFKF
                     VGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPGVETVGESASKESEL
                     KQSTEKQEGTLKQAQSSTEIPLQAESGQGTEEEAAKDGEENREKEPTKPLESPTSPVS
                     NETTSSFKKFFTHGWAGWRKKTSFKKPKEDDLETSEKRKEQEAEKVDEEEGEKTEPAP
                     AEEQEPAEGTDQARLSADYEKVELPLEDQVGDLEALSEKCAPLATEVFDEKTEAHQEV
                     VAEVHVSTVEKMTKGQGGAEVEGDVVVEGSGESLPPEKLAETQEVPQEAEPVEELMKT
                     KEVCVSGGDHTQLTDLSPEEKMLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSG
                     LKKLSGKKQKGKRGGGGDEEPGEYQHIQTESPESADEQKGESSASSPEEPEEIACLE
                     KGPSEAPQEAEAEEGATSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELDKV
                     KSATLSSTESTASGMQDEVRAVGEEQRSEEPKRRVDTSVSWEALICVGSSKKRARKAS
                     SSDDEGGPRTLGGDGHRAEEASKDKEADALPASTQEQDQAHGSSSPEPAGSPSEGEGV
                     STWESFKRLVTPRKKSKSKLEERAEDSGAEQLASEIEPSREESWVSIKKFIPGRRKKR
                     ADGKQEQAAVEDSGPGEINEDDPDVPAVVPLSEYDAVEREKLEAQRAQENVELPQLKG
                     AVYVSEELSKTLVHTVSVAVIDGTRAVTSAEERSPSWISASMTEPLEHAEGVATPPVG
                     EVTEKDITAEATPALAQTLPGGKDAHDDIVTSEVDFTSEAVTAAETTEALRAEELTEA
                     SGAEETTDMVSAVSQLSDSPDTTEEATPVQEVEGGMLDTEEQERQTQAVLQAVADKVK
                     EDSQVPATQTLQRAGPKALEKVEEVEEDSEVLATEKEKDVVPEGPVQEAETEHLAQGS
                     ETVQATPESLEVPEVTEDVDRATTCQVIKHQQLMEQAVAPESSETLTDSETNGSTPLA
                     DSDTPNGTQQDETVDSQDSNAIAAVKQSQVTEEEAAAAQTEGPSTPSSFPAQEEHREK
                     PGRDVLEPTQALAAGAVPILAKAEVGQEGEAGQFDGEKVKDGQCVKELEVPVHTGPNS
                     QKTADLTRDSEVMEVARCQETESNEEQSISPEKREMGTDVEKEETETKTEQASEEHEQ
                     ETAAPEHEGTHPKPVLTADMPHSERGKALGSLEGSPSLPDQDKADCIEVQVQSSDTPV
                     TQTTEAVKKVEETVATSEMDESLECAGAQSLPAEKLSETGGYGTLQHGEDTVPQGPES
                     QAESIPIIVTPAPESILHSDLQREVSASQKQRSDEDNKPDAGPDAAGKESAAREKILR
                     AEPEILELESKSNKIVQSVIQTAVDQFARTETAPETHASDLQNQVPVMQADSQGAQQM
                     LDKDESDLQVSPQDGTLSAVAQEGLAVSDSSEGMSKASEMITTLAVESASVKESVEKL
                     PLQCKDEKEHAADGPQHQSLAKAEADASGNLTKESPDTNGPKLTEEGDALKEEMNKAQ
                     TEEDDLQEPKGDLTES"
BASE COUNT     1840 a   1447 c   1779 g   1131 t
ORIGIN

  Query Match             74.0%;  Score 4559.6;  DB 10;  Length 6197;
  Best Local Similarity   86.6%;  Pred. No. 0;
  Matches 5375;  Conservative    0;  Mismatches  644;  Indels  189;  Gaps   24;

Qy        4 GCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGGA 63
Db      111 GCCGGCGAGAGTAGGAAAGCCACTGAGCCATGGGTGCAGGCAGTTCCACCGAGCAGCGGA 170

Qy       64 GCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATGGGC 123
Db      171 GCCCCGAGCAGCCGGCGGAGAGCGACACGCCGAGCGAGCTGGAGCTCAGTGGCCATGGGC 230

Qy      124 CCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCAAGC 183
Db      231 CCGCAGCGGAAGCGTCGGGAGCAGCTGGAGATCCCGCTGACGCGGACCCCGCCACCAAGC 290

Qy      184 TCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATGTCC 243
Db      291 TCCCACAGAAGAATGGTCAGCTGTCTGCCGTCAATGGTGTAGCTGAACAAGAAGATGTCC 350

Qy      244 ATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGAC 303
Db      351 ACGTCCAAGAGGAAAGCCAGGATGGGCAAGAGGAAGAAGTCACTGTTGAAGATGTTGGAC 410

Qy      304 AGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCA 363
Db      411 AGAGAGAGTCAGAAGATGTGAAAGAAAAAGACCGAGCTAAAGAAATGGCAGCCAGTTCCA 470

Qy      364 CAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGA 423
Db      471 CAGTTGTTGAAGATATCACAAAGGACGAGCAGGAGGAAACACCGGAAATAATCGAACAGA 530

Qy      424 TCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATG 483
Db      531 TCCCTGCTTCAGAGAGCAATGTGGAAGAAATGGCGCAGGCTGCTGAGTCCCAAGCTAATG 590

Qy      484 ATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATA 543
Db      591 ACGTCGGCTTCAAGAAGGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATA 650

Qy      544 AAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGG 603
Db      651 AAAACGAAAAGTCAGATACCGTCCAGCTACTCACTGTCAAGAAGGATGAAGGCGAAGGGG 710

Qy      604 CAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAG 663
Db      711 CAGAAGCCTCCGTCGGAGCAGGAGACCACCAAGAGCCCGGAGTGGAGA---CCGTCGGCG 767

Qy      664 AGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGA 723
Db      768 AATCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGA 827

Qy      724 AGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGG 783
Db      828 AGCAAGCACAGAGCAGCACAGAAATTCCCCTTCAAGCCGAATCTGGTCAAGGGACCGAGG 887

Qy      784 AAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAAT 843
Db      888 AAGAAGCAGCCAAAGATGGAGAAGAAAACCGAGAGAAAGAACCTACCAAGCCCCTAGAAT 947

Qy      844 CCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACG 903
Db      948 CTCCGACCAGCCCTGTCAGCAATGAGACAACATCTTCCTTCAAGAAATTCTTCACTCACG 1007

Qy      904 GTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAA 963
Db     1008 GGTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAACCAAAGGAAGATGATCTGGAAA 1067

Qy      964 CTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGA 1023
Db     1068 CTTCCGAGAAGAGAAAGGAGCAAGAGGCTGAAAAAGTAGACGAGGAAGAAGGGGAAAAGA 1127

Qy     1024 CAGAGCCAGCC---TCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
Db     1128 CAGAGCCAGCCCCAGCCGAGGAGCAGGAGCCTGCAGAAGGCACAGACCAGGCCAGGTTGT 1187

Qy     1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
Db     1188 CAGCCGACTATGAGAAGGTGGAGTTGCCTTTGGAAGACCAGGTCGGTGACCTGGAGGCAT 1247

Qy     1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
Db     1248 TGTC---GGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGACGGAAGCCC 1304

Qy     1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
```

```
Db   1305 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGATGACGAAAGGGCA-- 1362

Qy   1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
Db   1363 -AGGAGGAGCAGAGGTGGAAGGGGATGTGGTGGTGGAAGGATCGGGAGAATCCTTGCCCC 1421

Qy   1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
Db   1422 CTGAGAAACTGGCTGAGACCCAGGAGGTCCCCCAGGAAGCTGAGCCTGTGGAGGAGCTGA 1481

Qy   1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
Db   1482 TGAAGACCAAAGAAGTATGCGTCTCTGGGGGTGACCATACTCAGCTGACAGATCTAAGTC 1541

Qy   1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
Db   1542 CTGAAGAGAAGATGCTACCCAAACACCCCGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1601

Qy   1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
Db   1602 CCTCTCAGGAGAGAATCAAGGTACAGGGAAGTCCCCTGAAGAAGCTCTTCAGCAGTTCGG 1661

Qy   1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAA---ACGAGGAGGTGGGGGAGACG 1617
Db   1662 GCTTAAAGAAGCTCTCCGGGAAGAAGCAGAAGGGGAAGAGAGGAGGAGGCGGGGGAGATG 1721

Qy   1618 AAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGA 1677
Db   1722 AAGAGCCAGGAGAATACCAACACATTCAAACCGAGTCCCCAGAGAGTGCTGACGAGCAGA 1781

Qy   1678 AGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAG 1737
Db   1782 AGGGAGAGAGCTCTGCCTCTTCCCCTGAAGAGCCCGAGGAGATCGCGTGTCTGGAGAAGG 1841

Qy   1738 GGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGA 1797
Db   1842 GGCCATCGGAAGCACCCCAGGAAGCGGAAGCTGAGGAAGGAGCGACTTCCGACGGAGAGA 1901

Qy   1798 AGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAAC 1857
Db   1902 AGAAAAGGGAAGGGATCACCCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAAC 1961

Qy   1858 GGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCA 1917
Db   1962 GGGTCCGAAGACCTTCTGAGAGCGACAAGGAAGAAGAGCTGGATAAGGTCAAGAGTGCCA 2021

Qy   1918 CCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTG 1977
Db   2022 CCTTGTCCTCCACGGAGAGCACGGCGTCTGGAATGCAGGATGAGGTCAGAGCGGTTGGCG 2081

Qy   1978 AGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCAC 2037
Db   2082 AGGAGCAAAGGTCAGAGGAGCCAAAGCGCAGGGTGGATACTTCAGTGTCTTGGGAGGCGT 2141

Qy   2038 TGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAG 2097
Db   2142 TGATTTGTGTCGGATCGTCCAAGAAGAGAGCGAGGAAGGCATCCTCTTCAGATGATGAAG 2201

Qy   2098 GAGGGCCAAGGACACTGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAG 2157
Db   2202 GAGGGCCAAGAACACTGGAGGGGGATGGCCACAGAGCGGAGGAGGCTAGCAAAGACAAAG 2261

Qy   2158 AAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTT 2217
Db   2262 AAG------CAGATGCTCTTCCTGCCAGCACCCAGGAACAAGACCAAGCGCACGGAAGTT 2315

Qy   2218 CCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCAT 2277
Db   2316 CCTCACCCGAGCCAGCTGGAAGCCCTTCTGAAGGGGAGGCGTCTCCACCTGGGAGTCAT 2375

Qy   2278 TTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAG 2337
Db   2376 TTAAGAGATTAGTCACTCCACGAAAAAAATCCAAGTCAAAACTGGAAGAGAGAGCCGAAG 2435

Qy   2338 ACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTT 2397
Db   2436 ACTCCGGTGCAGAGCAGTTGGCCTCCGAGATCGAACCAAGTAGAGAGGAATCTTGGGTTT 2495

Qy   2398 CCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAG 2457
Db   2496 CCATTAAGAAATTTATTCCTGGACGGCGGAAGAAAAGGGCAGATGGAAGCAAGAACAGG 2555

Qy   2458 CCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCG 2517
Db   2556 CCGCCGTTGAAGACTCGGGGCCAGGAGAGATCAATGAGGACGACCCCGACGTCCCAGCTG 2615

Qy   2518 TCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGA---------AGCCC 2568
Db   2616 TTGTGCCTCTGTCTGAGTACGATGCGGTAGAGAGAGAGAAGCTGGAAGCGCAGCGAGCTC 2675

Qy   2569 AGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTA 2628
Db   2676 AGGAGAACGTGGAGCTGCCCCAGCTGAAGGGGGCTGTGTATGTGTCTGAGGAGCTTAGTA 2735

Qy   2629 AGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTG 2688
Db   2736 AGACTCTGGTTCACACTGTGAGTGTCGCGGTCATTGATGGGACCAGGGCAGTCACCAGTG 2795

Qy   2689 TCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAG 2748
Db   2796 CCGAAGAGCGGTCCCCTTCGTGGATATCTGCTTCCATGACAGAACCTCTTGAGCACGCAG 2855

Qy   2749 CGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAA 2808
Db   2856 AGGGAGTGGCCACACCGCCTGTTGGAGAGGTCACTGAAAAAGACATCACTGCAGAAGCAA 2915

Qy   2809 CTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCA 2868
Db   2916 CTCCTGCACTCGCCCAGACTTTACCAGGGGGCAAAGATGCCCATGACGACATAGTCACCA 2975

Qy   2869 GTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTA 2928
Db   2976 GTGAAGTGGATTTTACCTCAGAAGCAGTGACAGCCGCAGAAACCACAGAGGCGCTCCGCG 3035

Qy   2929 CTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTT 2988
Db   3036 CTGAAGAACTTACCGAAGCATCAGGGGCAGAAGAGACCACAGACATGGTGTCTGCAGTTT 3095

Qy   2989 CCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGA 3048
Db   3096 CCCAGCTGTCCGACTCCCCGGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGG 3155

Qy   3049 GTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTG 3108
Db   3156 GTGGCATGCTAGATACGGAAGAACAGGAGCGCCAGACGCAGGCCGTCCTCCAAGCCGTTG 3215

Qy   3109 CAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGT 3168
Db   3216 CAGACAAAGTGAAAGAGGACTCCCAGGTGCCTGCAACCCAGACTCTGCAGAGAGCAGGGC 3275

Qy   3169 CAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGA 3228
Db   3276 CGAAAGCACTGGAGAAGGTGGAGGAGGTAGAGGAGGACTCCGAGGTGCTGGCTACCGAGA 3335

Qy   3229 AAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCAC 3288
Db   3336 AAGAGAAGGATGTTGTGCCGGAAGGACCCGTGCAGGAAGCTGAAACTGAGCATCTTGCAC 3395

Qy   3289 AGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAG 3348
Db   3396 AGGGCTCCGAGACTGTACAGGCTACCCCAGAGAGCCTTGAAGTTCCTGAAGTCACAGAGG 3455

Qy   3349 ATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCG 3408
Db   3456 ATGTAGACCGTGCCACCACATGCCAGGTTATCAAGCACCAGCAGCTGATGGAACAGGCTG 3515
```

```
Qy   3409 TGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAG 3468
Db   3516 TGGCCCCTGAGTCATCTGAAACCTTGACAGACAGTGAGACAAATGGAAGTACTCCCCTCG 3575

Qy   3469 CAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTA 3528
Db   3576 CAGATTCAGACACTCCAAACGGGACACAGCAAGACGAGACCGTTGACAGCCAGGACAGTA 3635

Qy   3529 AAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGA 3588
Db   3636 ATGCCATTGCCGCCGTCAAGCAGTCACAGGTCACTGAAGAGGAGGCAGCTGCTGCTCAGA 3695

Qy   3589 AAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAAC 3648
Db   3696 CGGAGGGGCCTTCAACACCATC---TAGTTTTCCAGCCCAGGAAGAACACAGGGAAAAAC 3752

Qy   3649 CAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTC 3708
Db   3753 CAGGAAGGGATGTTCTAGAACCCACA---CAAGCGCTGGCTGCCGGGGCAGTGCCTATTC 3809

Qy   3709 TGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCA 3768
Db   3810 TGGCAAAGGCTGAGGTGGGTCAAGAGGGTGAGGCTGGCCAGTTTGATGGAGAAAAAGTCA 3869

Qy   3769 AAG--------------AAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAA 3813
Db   3870 AAGACGGACAGTGTGTTAAAGAACTGGAGGTGCCTGTGCACACTGGACCCAACAGTCAAA 3929

Qy   3814 AGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGG 3873
Db   3930 AGACTGCTGACTTGACACGTGACAGTGAAGTAATGGAAGTGGCCAGATGTCAGGAAACTG 3989

Qy   3874 AGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAA 3933
Db   3990 AGAGTAATGAAGAACAGAGTATTAGCCCGGAGAAAAGAGAGATGGAACCGACGTTGAAA 4049

Qy   3934 AGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCG 3993
Db   4050 AGGAGGAAACAGAGACCAAGACAGAGCAAGCCAGTGAAGAACATGAGCAGGAAACAGCTG 4109

Qy   3994 CTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAG 4053
Db   4110 CTCCTGAGCATGAAGGAACCCACCCTAAGCCAGTCCTGACAGCTGACATGCCTCACTCAG 4169

Qy   4054 AGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAG 4113
Db   4170 AGAGGGGAAAGGCACTGGGCAGCCTTGAAGGAAGCCCTTCTCTCCCAGACCAAGACAAAG 4229

Qy   4114 CAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAG 4173
Db   4230 CAGATTGCATAGAGGTTCAAGTTCAAAGCTCAGACACACCAGTCACTCAAACAACCGAAG 4289

Qy   4174 CTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTG 4233
Db   4290 CTGTGAAAAAGGTCGAAGAAACTGTGGCAACTTCAGAGATGGATGAAAGTTTGGAGTGTG 4349

Qy   4234 TAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTC 4293
Db   4350 CAGGTGCGCAATCATTACCAGCTGAGAAGCTCTCCGAAACCGGTGGCTACGGGACTCTTC 4409

Qy   4294 AGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCA 4353
Db   4410 AGCATGGAGAGGACACCGTGCCCCAGGGGCCTGAGTCTCAGGCAGAGTCCATCCCAATAA 4469

Qy   4354 TAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCAT 4413
Db   4470 TAGTAACTCCTGCTCCTGAAAGCATCCTACATTCTGACCTTCAAAGAGAAGTGAGCGCAT 4529

Qy   4414 CCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCA 4473
Db   4530 CCCAGAAACAGAGATCAGATGAAGATAACAAGCCAGATGCTGGTCCTGATGCTGCCGGCA 4589

Qy   4474 AGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGA 4533
Db   4590 AGGAGAGTGCAGCAAGAGAGAAAATCCTCAGGGCTGAACCTGAGATCTTGGAACTTGAGA 4649

Qy   4534 GTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTA 4593
Db   4650 GTAAGAGCAATAAGATTGTCCAGAGTGTCATCCAGACAGCCGTCGACCAGTTTGCACGTA 4709

Qy   4594 CAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTG-CATGCAGG 4652
Db   4710 CAGAAACAGCCCCCGAAACCCACGCTTCTGATTTACAGAATCAGGTTCCTGTGATGCAGG 4769

Qy   4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAG--------------AT 4697
Db   4770 CTGACAGCCAGGGAGCACAACAGATGCTGGACAAAGATGAAAGCGACCTTCAAGTCTCCC 4829

Qy   4698 GCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTG 4757
Db   4830 CCCAAGATGGAACACTCAGTGCCGTAGCCCAGGAAGGACTTGCGGTTTCTGATAGTTCTG 4889

Qy   4758 GAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCA 4817
Db   4890 AAGGCATGAGCAAGGCTTCAGAAATGATCACCACGCTTGCAGTTGAAAGTGCCAGTGTCA 4949

Qy   4818 AAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGAT 4877
Db   4950 AAGAAAGTGTTGAGAAGCTGCCTCTTCAGTGCAAAGAT-GAAAAGGAACATGCTGCTGAC 5008

Qy   4878 GGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAA 4937
Db   5009 GGCCCCCAGCACCAAAGCTTAGCCAAGGCAGAGGCGGATGCCTCTGGAAATCTAACCAAA 5068

Qy   4938 GAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAG 4997
Db   5069 GAGTCCCCAGACACCAACGGACCAAAGCTAACCGAGGAAGGAGATGCCCTGAAAGAAGAA 5128

Qy   4998 GTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAG 5057
Db   5129 AT-----------------------------------GAACAAGGCCCAGACAGAAGAG 5152

Qy   5058 ---GACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCATTG 5114
Db   5153 GACGACCTACAGGAGCCAAAGGGAGACCTGACAGAATCCTAAGACGTGAGTTGCTCAT-- 5210

Qy   5115 TACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACC 5174
Db   5211 ------TGTAAAACTAGAATGTGAAGTGAAGTCACAGAACAAGATGCTGCTGTTGGGACC 5264

Qy   5175 TTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCCA 5234
Db   5265 TTGAGACCAAAATTTCAGAGCCCTTGAGGTGCAGAGAGCAGAGCCGTCCAATGATTTCAA 5324

Qy   5235 CCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATTT 5294
Db   5325 CCCTGCAGAGCACCCCGACAATCCTGAGGCTGCATCGGGAGCTAGAGCCAGCTAACATTT 5384

Qy   5295 CCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATTT 5354
Db   5385 CCTCTTTTCAAGACTGCCCTTGATTTGCCCCTTGGTG---------------------- 5421

Qy   5355 CCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTAA 5414
Db   5422 ----------------------------------------CCATGTATTTCAGAGTTAA 5440

Qy   5415 GGTCCTGCGTTCTCAACCTGGAACCAAT-TCTGCCATACCTAGTTCCACTTCTCAAACTG 5473
Db   5441 GGTCCTGCTTTCTCCACCTGGAACCAATGTCGGCAATACCTAGTCCCACTTCTCAAACTG 5500

Qy   5474 GAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATT 5533
Db   5501 GAGCCTCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATT 5560

Qy   5534 GTATA-TTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACGGGT 5592
```

```
Db     5561 GTATATTTTTTTCTAAGGTTTAAGCACATGCCTTTTGTATTATGCAGTATATAATGGGT 5620

Qy     5593 GTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAG 5652
Db     5621 GTGCAGCCATAGCGAAGCTTTGAGAAGCCCCAAGCCTCAACTGTAACCTGCAGCAAACAG 5680

Qy     5653 ----ATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTAAAGTTTACTGATGCTTAGAT 5708
Db     5681 AGAAATAACATTCCTGGCAGGAAGATACAAGTC-TTTTTAAAGTTTACTGATGCTTAACT 5739

Qy     5709 CTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATA 5768
Db     5740 CTGTGGGCTCGTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGTGAGCTCAGAAATA 5799

Qy     5769 AAAACCCCATTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTT 5828
Db     5800 AAAACTCCATTTTGAGACATTCAGAATGTCCCAACATTACGACAACATTTTTTTTTTTT 5859

Qy     5829 TTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCT 5888
Db     5860 TTTCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTTAGTGTCAGATTAAGCCCCATCTCT 5919

Qy     5889 TAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAA 5948
Db     5920 TAACAATATGGACAGATGAGTGTGCC-ATGGCCATGAGCTATTTCCTAACGCAGAAGGAA 5978

Qy     5949 TCTGTTGTACGTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCG 6008
Db     5979 TTTGTTGTTACTTTTT-GGATTGTACTCTTCTAGGCTGGACCGAATTCATATGCAGATTG 6037

Qy     6009 AAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTAT 6068
Db     6038 AAGTGAGACCTGCTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTATTAT 6097

Qy     6069 ATCTGTGCCCCTTC-TTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATT 6127
Db     6098 ATCTGTGCCCCTTCTTTTAAGAACAATGTTACATTATGTTCCTTTGGATAAATTGTGATT 6157

Qy     6128 TGACAACTGATTTAAATAAACATATTTG 6155
Db     6158 TGACAACTGATTTCAATAAAAATATTTG 6185
```

```
RESULT 5
AB020886
LOCUS       AB020886                6195 bp    mRNA    linear   ROD 14-APR-1999
DEFINITION  Mus musculus mRNA for SSeCKS, complete cds.
ACCESSION   AB020886
VERSION     AB020886.1  GI:4589831
KEYWORDS    SSeCKS.
SOURCE      Mus musculus (strain:BALB/c, isolate:200) 9-12 weeks male testis
            tissue_lib:Clontech mouse testis library cDNA to mRNA, clone:133.5.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Okita,K., Kitamura,H., Iwanaga,T., Morimatsu,M. and Saito,M.
  TITLE     Isolation and characterization of mouse SSeCKS
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6195)
  AUTHORS   Okita,K., Kitamura,H., Saito,M. and Morimatsu,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1998) Keisuke Okita, School of Veterinary
            Medicine, Hokkaido University, Laboratory of Biochemistry; N18, W9,
            Kita-ku, Sapporo, Hokkaido 060-0818, Japan
            (E-mail:v080008@vet.ec.hokudai.ac.jp, Tel:81-11-706-5206,
            Fax:81-11-757-0703)
COMMENT     Sequence updated (10-Mar-1999).
FEATURES             Location/Qualifiers
     source          1..6195
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /isolate="200"
                     /db_xref="taxon:10090"
                     /clone="133.5"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="9-12 weeks"
                     /tissue_lib="Clontech mouse testis library"
     gene            1..6195
                     /gene="ssecks"
     CDS             140..5194
                     /gene="ssecks"
                     /note="Src Suppressed C Kinase Substrate"
                     /codon_start=1
                     /product="SSeCKS"
                     /protein_id="BAA76894.1"
                     /db_xref="GI:4589832"
                     /translation="MGAGSSTEQRSPEQPAESDTPSELELSGHGPAAEASGAAGDPAD
                     ADPATKLPQKNGQLSAVNGVAEQEDVHVQEESQDGQEEEVTVEDVGQRESEDVKEKDR
                     AKEMAASSTVVEDITKDEQEETPEIIEQIPASESNVEEMAQAAESQANDVGFKKVFKF
                     VGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPGVETVGESASKESEL
                     KQSTEKQEGTLKQAQSSTEIPLQAESGQGTEEEAAKDGEENREKEPTKPLESPTSPVS
                     NETTSSFKKFFTHGWAGWRKKTSFKKPKEDDLETSEKRKEQEAEKVDEEEGEKTEPAP
                     AEEQEPAEGTDQARLSADYEKVELPLEDQVGDLEALSEKCAPLATEVFDEKTEAHQEV
                     VAEVHVSTVEKMTKGQGGAEVEGDVVVEGSGESLPPEKLAETQEVPQEAEPVEELMKT
                     KEVCVSGGDHTQLTDLSPEEKMLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSG
                     LKKLSGKKQKGKRGGGGDEEPGEYQHIQTESPESADEQKGESSASSPEEPEEIACLE
                     KGPSEAPQEAEAEEGATSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELDKV
                     KSATLSSTESTASGMQDEVRAVGEEQRSEEPKRRVDTSVSWEALICVGSSKKRARKAS
                     SSDDEGGPRTLGGDGHRAEEASKDKEADALPASTQEQDQAHGSSSPEPAGSPSEGEGV
                     STWESFKRLVTPRKKSKSKLEERAEDSGAEQLASEIEPSREESWVSIKKFIPGRRKKR
                     ADGKQEQAAVEDSGPGEINEDDPDVPAVVPLSEYDAVEREKLEAQRAQENVELPQLKG
                     AVYVSEELSKTLVHTVSVAVIDGTRAVTSAEERSPSWISASMTEPLEHAEGVATPPVG
                     EVTEKDITAEATPALAQTLPGGKDAHDDIVTSEVDFTSEAVTAAETTEALRAEELTEA
                     SGAEETTDMVSAVSQLSDSPDTTEEATPVQEVEGGMLDTEEQERQTQAVLQAVADKVK
                     EDSQVPATQTLQRAGPKALEKVEEVEEDSEVLATEKEKDVVPEGPVQEAETEHLAQGS
                     ETVQATPESLEVPEVTEDVDRATTCQVIKHQQLMEQAVAPESSETLTDSETNGSTPLA
                     DSDTPNGTQQDETVDSQDSNAIAAVKQSQVTEEEAAAAQTEGPSTPSSFPAQEEHREK
                     PGRDVLEPTQALAAGAVPILAKAEVGQEGEAGQFDGEKVKDGQCVKELEVPVHTGPNS
                     QKTADLTRDSEVMEVARCQETESNEEQSISPEKREMGTDVEKEETETKTEQASEEHEQ
                     ETAAPEHEGTHPKPVLTADMPHSERGKALGSLEGSPSLPDQDKADCIEVQVQSSDTPV
                     TQTTEAVKKVEETVATSEMDESLECAGAQSLPAEKLSETGGYGTLQHGEDTVPQGPES
                     QAESIPIIVTPAPESILHSDLQREVSASQKQRSDEDNKPDAGPDAAGKESAAREKILR
                     AEPEILELESKSNKIVQSVIQTAVDQFARTETAPETHASDLQNQVPVMQADSQGAQQM
                     LDKDESDLQVSPQDGTLSAVAQEGLAVSDSSEGMSKASEMITTLAVESASVKESVEKL
                     PLQCKDEKEHAADGPQHQSLAKAEADASGNLTKESPDTNGPKLTEEGDALKEEMNKAQ
                     TEEDDLQEPKGDLTES"
     polyA_site      6195
                     /gene="ssecks"
                     /note="20 A nucleotides"
BASE COUNT     1838 a   1451 c   1777 g   1129 t
ORIGIN

  Query Match             74.0%;  Score 4558;  DB 10;  Length 6195;
  Best Local Similarity   86.6%;  Pred. No. 0;
  Matches 5374;  Conservative    0;  Mismatches  645;  Indels  189;  Gaps   24;

Qy        4 GCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGGA 63
Db      111 GCCGGCGAGAGTAGGAAAGCCACTGAGCCATGGGTGCAGGCAGTTCCACCGAGCAGCGGA 170

Qy       64 GCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATGGGC 123
Db      171 GCCCCGAGCAGCCGGCGGAGAGCGACACGCCGAGCGAGCTGGAGCTCAGTGGCCATGGGC 230

Qy      124 CCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCAAGC 183
Db      231 CCGCAGCGGAAGCGTCGGGAGCAGCTGGAGATCCCGCTGACGCGGACCCCGCCACCAAGC 290

Qy      184 TCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATGTCC 243
Db      291 TCCCACAGAAGAATGGTCAGCTGTCTGCCGTCAATGGTGTAGCTGAACAAGAAGATGTCC 350

Qy      244 ATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGAC 303
```

```
Db     351 ACGTCCAAGAGGAAAGCCAGGATGGGCAAGAGGAAGAAGTCACTGTTGAAGATGTTGGAC 410

Qy     304 AGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCA 363
Db     411 AGAGAGAGTCAGAAGATGTGAAAGAAAAAGACCGAGCTAAAGAAATGGCAGCCAGTTCCA 470

Qy     364 CAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGA 423
Db     471 CAGTTGTTGAAGATATCACAAAGGACGAGCAGGAGGAAACACCGGAAATAATCGAACAGA 530

Qy     424 TCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATG 483
Db     531 TCCCTGCTTCAGAGAGCAATGTGGAAGAAATGGCGCAGGCTGCTGAGTCCCAAGCTAATG 590

Qy     484 ATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATA 543
Db     591 ACGTCGGCTTCAAGAAGGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATA 650

Qy     544 AAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGG 603
Db     651 AAAACGAAAAGTCAGATACCGTCCAGCTACTCACTGTCAAGAAGGATGAAGGCGAAGGGG 710

Qy     604 CAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAG 663
Db     711 CAGAAGCCTCCGTCGGAGCAGGAGACCACCAAGAGCCCGGAGTGGAGA---CCGTCGGCG 767

Qy     664 AGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGA 723
Db     768 AATCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGA 827

Qy     724 AGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGG 783
Db     828 AGCAAGCACAGAGCAGCACAGAAATTCCCCTTCAAGCCGAATCTGGTCAAGGGACCGAGG 887

Qy     784 AAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAAT 843
Db     888 AAGAAGCAGCCAAAGATGGAGAAGAAAACCGAGAGAAAGAACCTACCAAGCCCCTAGAAT 947

Qy     844 CCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACG 903
Db     948 CTCCGACCAGCCCTGTCAGCAATGAGACAACATCTTCCTTCAAGAAATTCTTCACTCACG 1007

Qy     904 GTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAA 963
Db    1008 GCTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAACCAAAGGAAGATGATCTGGAAA 1067

Qy     964 CTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGA 1023
Db    1068 CTTCCGAGAAGAGAAAGGAGCAAGAGGCTGAAAAAGTAGACGAGGAAGAAGGGGAAAAGA 1127

Qy    1024 CAGAGCCAGCC---TCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
Db    1128 CAGAGCCAGCCCCAGCCGAGGAGCAGGAGCCTGCAGAAGGCACAGACCAGGCCAGGTTGT 1187

Qy    1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
Db    1188 CAGCCGACTATGAGAAGGTGGAGTTGCCTTTGGAAGACCAGGTCGGTGACCTGGAGGCAT 1247

Qy    1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
Db    1248 TGTC---GGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGACGGAAGCCC 1304

Qy    1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
Db    1305 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGATGACGAAAGGGCA-- 1362

Qy    1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
Db    1363 -AGGAGGAGCAGAGGTGGAAGGGGATGTGGTGGTGGAAGGATCGGGAGAATCCTTGCCCC 1421

Qy    1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
Db    1422 CTGAGAAACTGGCTGAGACCCAGGAGGTCCCCCAGGAAGCTGAGCCTGTGGAGGAGCTGA 1481

Qy    1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
Db    1482 TGAAGACCAAAGAAGTATGCGTCTCTGGGGGTGACCATACTCAGCTGACAGATCTAAGTC 1541

Qy    1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
Db    1542 CTGAAGAGAAGATGCTACCCAAACACCCCGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1601

Qy    1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
Db    1602 CCTCTCAGGAGAGAATCAAGGTACAGGGAAGTCCCCTGAAGAAGCTCTTCAGCAGTTCGG 1661

Qy    1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAA---ACGAGGAGGTGGGGGAGACG 1617
Db    1662 GCTTAAAGAAGCTCTCCGGGAAGAAGCAGAAGGGGAAGAGAGGAGGAGGCGGGGGAGATG 1721

Qy    1618 AAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGA 1677
Db    1722 AAGAGCCAGGAGAATACCAACACATTCAAACCGAGTCCCCAGAGAGTGCTGACGAGCAGA 1781

Qy    1678 AGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAG 1737
Db    1782 AGGGAGAGAGCTCTGCCTCTTCCCCTGAAGAGCCCGAGGAGATCGCGTGTCTGGAGAAGG 1841

Qy    1738 GGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGA 1797
Db    1842 GGCCATCGGAAGCACCCCAGGAAGCGGAAGCTGAGGAAGGAGCGACTTCCGACGGAGAGA 1901

Qy    1798 AGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAAC 1857
Db    1902 AGAAAAGGGAAGGGATCACCCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAAC 1961

Qy    1858 GGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCA 1917
Db    1962 GGGTCCGAAGACCTTCTGAGAGCGACAAGGAAGAAGAGCTGGATAAGGTCAAGAGTGCCA 2021

Qy    1918 CCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTG 1977
Db    2022 CCTTGTCCTCCACGGAGAGCACGGCGTCTGGAATGCAGGATGAGGTCAGAGCGGTTGGCG 2081

Qy    1978 AGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCAC 2037
Db    2082 AGGAGCAAAGGTCAGAGGAGCCAAAGCGCAGGGTGGATACTTCAGTGTCTTGGGAGGCGT 2141

Qy    2038 TGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAG 2097
Db    2142 TGATTTGTGTCGGATCGTCCAAGAAGAGAGCGAGGAAGGCATCCTCTTCAGATGATGAAG 2201

Qy    2098 GAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAG 2157
Db    2202 GAGGGCCAAGAACACTGGGAGGGGATGGCCACAGAGCGGAGGAGGCTAGCAAAGACAAAG 2261

Qy    2158 AAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTT 2217
Db    2262 AAG------CAGATGCTCTTCCTGCCAGCACCCAGGAACAAGACCAAGCGCACGGAAGTT 2315

Qy    2218 CCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCAT 2277
Db    2316 CCTCACCCGAGCCAGCTGGAAGCCCTTCTGAAGGGGAGGGCGTCTCCACCTGGGAGTCAT 2375

Qy    2278 TTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAG 2337
Db    2376 TTAAGAGATTAGTCACTCCACGAAAAAAATCCAAGTCAAAACTGGAAGAGAGAGCCGAAG 2435

Qy    2338 ACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTT 2397
Db    2436 ACTCCGGTGCAGAGCAGTTGGCCTCCGAGATCGAACCAAGTAGAGAGGAATCTTGGGTTT 2495

Qy    2398 CCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAG 2457
Db    2496 CCATTAAGAAATTTATTCCTGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAGAACAGG 2555
```

```
Qy   2458 CCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCG 2517
Db   2556 CCGCCGTTGAAGACTCGGGGCCAGGAGAGATCAATGAGGACGACCCCGACGTCCCAGCTG 2615

Qy   2518 TCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGA---------AGCCC 2568
Db   2616 TTGTGCCTCTGTCTGAGTACGATGCGGTAGAGAGAGAGAAGCTGGAAGCGCAGCGAGCTC 2675

Qy   2569 AGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTA 2628
Db   2676 AGGAGAACGTGGAGCTGCCCCAGCTGAAGGGGGCTGTGTATGTGTCTGAGGAGCTTAGTA 2735

Qy   2629 AGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTG 2688
Db   2736 AGACTCTGGTTCACACCGTGAGTGTCGCGGTCATTGATGGGACCAGGGCAGTCACCAGTG 2795

Qy   2689 TCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAG 2748
Db   2796 CCGAAGAGCGGTCCCCTTCGTGGATATCTGCTTCCATGACAGAACCTCTTGAGCACGCAG 2855

Qy   2749 CGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAA 2808
Db   2856 AGGGAGTGGCCACACCGCCTGTTGGAGAGGTCACTGAAAAAGACATCACTGCAGAAGCAA 2915

Qy   2809 CTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCA 2868
Db   2916 CTCCTGCACTCGCCCAGACTTTACCAGGGGGCAAAGATGCCCATGACGACATAGTCACCA 2975

Qy   2869 GTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTA 2928
Db   2976 GTGAAGTGGATTTTACCTCAGAAGCAGTGACAGCCGCAGAAACCACAGAGGCGCTCCGCG 3035

Qy   2929 CTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTT 2988
Db   3036 CTGAAGAACTTACCGAAGCATCAGGGGCAGAAGAGACCACAGACATGGTGTCTGCAGTTT 3095

Qy   2989 CCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGA 3048
Db   3096 CCCAGCTGTCCGACTCCCCGGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGG 3155

Qy   3049 GTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTG 3108
Db   3156 GTGGCATGCTAGATACGGAAGAACAGGAGCGCCAGACGCAGGCCGTCCTCCAAGCCGTTG 3215

Qy   3109 CAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGT 3168
Db   3216 CAGACAAAGTGAAAGAGGACTCCCAGGTGCCTGCAACCCAGACTCTGCAGAGAGCAGGGC 3275

Qy   3169 CAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGA 3228
Db   3276 CGAAAGCACTGGAGAAGGTGGAGGAGGTAGAGGAGGACTCCGAGGTGCTGGCTACCGAGA 3335

Qy   3229 AAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCAC 3288
Db   3336 AAGAGAAGGATGTTGTGCCGGAAGGACCCGTGCAGGAAGCTGAAACTGAGCATCTTGCAC 3395

Qy   3289 AGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAG 3348
Db   3396 AGGGCTCCGAGACTGTACAGGCTACCCCAGAGAGCCTTGAAGTTCCTGAAGTCACAGAGG 3455

Qy   3349 ATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCG 3408
Db   3456 ATGTAGACCGTGCCACCACATGCCAGGTTATCAAGCACCAGCAGCTGATGGAACAGGCTG 3515

Qy   3409 TGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAG 3468
Db   3516 TGGCCCCTGAGTCATCTGAAACCTTGACAGACAGTGAGACAAATGGAAGTACTCCCCTCG 3575

Qy   3469 CAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTA 3528
Db   3576 CAGATTCAGACACTCCAAACGGGACACAGCAAGACGAGACCGTTGACAGCCAGGACAGTA 3635

Qy   3529 AAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGA 3588
Db   3636 ATGCCATTGCCGCCGTCAAGCAGTCACAGGTCACTGAAGAGGAGGCAGCTGCTGCTCAGA 3695

Qy   3589 AAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAAC 3648
Db   3696 CGGAGGGGCCTTCAACACCATC---TAGTTTCCAGCCCAGGAAGAACACAGGGAAAAAC 3752

Qy   3649 CAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTC 3708
Db   3753 CAGGAAGGGATGTTCTAGAACCCACA---CAAGCGCTGGCTGCCGGGGCAGTGCCTATTC 3809

Qy   3709 TGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCA 3768
Db   3810 TGGCAAAGGCTGAGGTGGGTCAAGAGGGTGAGGCTGGCCAGTTTGATGGAGAAAAAGTCA 3869

Qy   3769 AAG---------------AAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAA 3813
Db   3870 AAGACGGACAGTGTGTTAAAGAACTGGAGGTGCCTGTGCACACTGGACCCAACAGTCAAA 3929

Qy   3814 AGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGG 3873
Db   3930 AGACTGCTGACTTGACACGTGACAGTGAAGTAATGGAAGTGGCCAGATGTCAGGAAACTG 3989

Qy   3874 AGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAA 3933
Db   3990 AGAGTAATGAAGAACAGAGTATTAGCCCGGAGAAAAGAGAGATGGGAACCGACGTTGAAA 4049

Qy   3934 AGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCG 3993
Db   4050 AGGAGGAAACAGAGACCAAGACAGAGCAAGCCAGTGAAGAACATGAGCAGGAAACAGCTG 4109

Qy   3994 CTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAG 4053
Db   4110 CTCCTGAGCATGAAGGAACCCACCCTAAGCCAGTCCTGACAGCTGACATGCCTCACTCAG 4169

Qy   4054 AGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAG 4113
Db   4170 AGAGGGGAAAGGCACTGGGCAGCCTTGAAGGAAGCCCTTCTCTCCCAGACCAAGACAAAG 4229

Qy   4114 CAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAG 4173
Db   4230 CAGATTGCATAGAGGTTCAAGTTCAAAGCTCAGACACACCAGTCACTCAAACAACCGAAG 4289

Qy   4174 CTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTG 4233
Db   4290 CTGTGAAAAAGGTCGAAGAAACTGTGGCAACTTCAGAGATGGATGAAAGTTTGGAGTGTG 4349

Qy   4234 TAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTC 4293
Db   4350 CAGGTGCGCAATCATTACCAGCTGAGAAGCTCTCCGAAACCGGTGGCTACGGGACTCTTC 4409

Qy   4294 AGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCA 4353
Db   4410 AGCATGGAGAGGACACCGTGCCCCAGGGGCCTGAGTCTCAGGCAGAGTCCATCCCAATAA 4469

Qy   4354 TAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCAT 4413
Db   4470 TAGTAACTCCTGCTCCTGAAAGCATCCTACATTCTGACCTTCAAAGAGAAGTGAGCGCAT 4529

Qy   4414 CCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCA 4473
Db   4530 CCCAGAAACAGAGATCAGATGAAGATAACAAGCCAGATGCTGGTCCTGATGCTGCCGGCA 4589

Qy   4474 AGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGA 4533
Db   4590 AGGAGAGTGCAGCAAGAGAGAAAATCCTCAGGGCTGAACCTGAGATCTTGGAACTTGAGA 4649

Qy   4534 GTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTA 4593
Db   4650 GTAAGAGCAATAAGATTGTCCAGAGTGTCATCCAGACAGCCGTCGACCAGTTTGCACGTA 4709

Qy   4594 CAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTG-CATGCAGG 4652
```

```
Db    4710 CAGAAACAGCCCCCGAAACCCACGCTTCTGATTTACAGAATCAGGTTCCTGTGATGCAGG 4769

Qy    4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAG--------------AT 4697

Db    4770 CTGACAGCCAGGGAGCACAACAGATGCTGGACAAAGATGAAAGCGACCTTCAAGTCTCCC 4829

Qy    4698 GCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTG 4757

Db    4830 CCCAAGATGGAACACTCAGTGCCGTAGCCCAGGAAGGACTTGCGGTTTCTGATAGTTCTG 4889

Qy    4758 GAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCA 4817

Db    4890 AAGGCATGAGCAAGGCTTCAGAAATGATCACCACGCTTGCAGTTGAAAGTGCCAGTGTCA 4949

Qy    4818 AAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGAT 4877

Db    4950 AAGAAAGTGTTGAGAAGCTGCCTCTTCAGTGCAAAGAT-GAAAAGGAACATGCTGCTGAC 5008

Qy    4878 GGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAA 4937

Db    5009 GGCCCCCAGCACCAAAGCTTAGCCAAGGCAGAGGCGGATGCCTCTGGAAATCTAACCAAA 5068

Qy    4938 GAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAG 4997

Db    5069 GAGTCCCCAGACACCAACGGACCAAAGCTAACCGAGGAAGGAGATGCCCTGAAAGAAGAA 5128

Qy    4998 GTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAG 5057

Db    5129 AT----------------------------------GAACAAGGCCCAGACAGAAGAG 5152

Qy    5058 ---GACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCATTG 5114

Db    5153 GACGACCTACAGGAGCCAAAGGGAGACCTGACAGAATCCTAAGACGTGAGTTGCTCAT-- 5210

Qy    5115 TACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACC 5174

Db    5211 ------TGTAAAACTAGAATGTGAAGTGAAGTCACAGAACAAGATGCTGCTGTTGGGACC 5264

Qy    5175 TTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCCA 5234

Db    5265 TTGAGACCAAAATTTCAGAGCCCTTGAGGTGCAGAGAGCAGAGCCGTCCAATGATTTCAA 5324

Qy    5235 CCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATTT 5294

Db    5325 CCCTGCAGAGCACCCCGACAATCCTGAGGCTGCATCGGGAGCTAGAGCCAGCTAACATTT 5384

Qy    5295 CCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATTT 5354

Db    5385 CCTCTTTTCAAGACTGCCCTTGATTTGCCCCTTGGTG---------------------- 5421

Qy    5355 CCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTAA 5414

Db    5422 -----------------------------------------CCATGTATTTCAGAGTTAA 5440

Qy    5415 GGTCCTGCGTTCTCAACCTGGAACCAAT-TCTGCCATACCTAGTTCCACTTCTCAAACTG 5473

Db    5441 GGTCCTGCTTTCTCCACCTGGAACCAATGTCGGCAATACCTAGTCCCACTTCTCAAACTG 5500

Qy    5474 GAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATT 5533

Db    5501 GAGCCTCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATT 5560

Qy    5534 GTATA-TTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACGGGT 5592

Db    5561 GTATATTTTTTTTCTAAGGTTTAAGCACATGCCTTTTGTATTATGCAGTATATAATGGGT 5620

Qy    5593 GTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAG 5652

Db    5621 GTGCAGCCATAGCGAAGCTTTGAGAAGCCCCAAGCCTCAACTGTAACCTGCAGCAAACAG 5680

Qy    5653 ----ATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGAT 5708

Db    5681 AGAAATAACATTCCTGGCAGGAAGATACAAGTC-TTTTTAAAGTTTACTGATGCTTAACT 5739

Qy    5709 CTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATA 5768

Db    5740 CTGTGGGCTCGTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGTGAGCTCAGAAATA 5799

Qy    5769 AAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTT 5828

Db    5800 AAAACTCCATTTTGAGACATTCAGAATGTCCCAACATTACGACAACATTTTTTTTTTTTT 5859

Qy    5829 TTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCT 5888

Db    5860 TTTCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTTAGTGTCAGATTAAGCCCCATCTCT 5919

Qy    5889 TAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAA 5948

Db    5920 TAACAATATGGACAGATGAGTGTGCC-ATGGCCATGAGCTATTTCCTAACGCAGAAGGAA 5978

Qy    5949 TCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCG 6008

Db    5979 TTTGTTGTTACTTTTT-GGATTGTACTCTTCTAGGCTGGACCGAATTCATATGCAGATTG 6037

Qy    6009 AAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTAT 6068

Db    6038 AAGTGAGACCTGCTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTATTAT 6097

Qy    6069 ATCTGTGCCCCTTC-TTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATT 6127

Db    6098 ATCTGTGCCCCTTCTTTTAAGAACAATGTTACATTATGTTCCTTTGGATAAATTGTGATT 6157

Qy    6128 TGACAACTGATTTAAATAAACATATTTG 6155

Db    6158 TGACAACTGATTTCAATAAAAATATTTG 6185
```

```
RESULT 6
AC115647
LOCUS       AC115647              178954 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-171G16, *** SEQUENCING IN PROGRESS
            ***, 53 unordered pieces.
ACCESSION   AC115647
VERSION     AC115647.3  GI:21737529
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 178954)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
```

```
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178954)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 178954)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:21239989.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GPHP
                Center clone name: CH230-171G16
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 142023 bases at least Q40
                Consensus quality: 146696 bases at least Q30
                Consensus quality: 150215 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 53 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *          1      1459: contig of 1459 bp in length
            *       1460      1559: gap of unknown length
            *       1560      2931: contig of 1372 bp in length
            *       2932      3031: gap of unknown length
            *       3032      4420: contig of 1389 bp in length
            *       4421      4520: gap of unknown length
            *       4521      5800: contig of 1280 bp in length
            *       5801      5900: gap of unknown length
            *       5901      7559: contig of 1659 bp in length
            *       7560      7659: gap of unknown length
            *       7660      8887: contig of 1228 bp in length
            *       8888      8987: gap of unknown length
            *       8988     10297: contig of 1310 bp in length
            *      10298     10397: gap of unknown length
            *      10398     12090: contig of 1693 bp in length
            *      12091     12190: gap of unknown length
            *      12191     13522: contig of 1332 bp in length
            *      13523     13622: gap of unknown length
            *      13623     15629: contig of 2007 bp in length
            *      15630     15729: gap of unknown length
            *      15730     17540: contig of 1811 bp in length
            *      17541     17640: gap of unknown length
            *      17641     19819: contig of 2179 bp in length
            *      19820     19919: gap of unknown length
            *      19920     21261: contig of 1342 bp in length
            *      21262     21361: gap of unknown length
            *      21362     23785: contig of 2424 bp in length
            *      23786     23885: gap of unknown length
            *      23886     25562: contig of 1677 bp in length
            *      25563     25662: gap of unknown length
            *      25663     27442: contig of 1780 bp in length
            *      27443     27542: gap of unknown length
            *      27543     30415: contig of 2873 bp in length
            *      30416     30515: gap of unknown length
            *      30516     31784: contig of 1269 bp in length
            *      31785     31884: gap of unknown length
            *      31885     34970: contig of 3086 bp in length
            *      34971     35070: gap of unknown length
            *      35071     36233: contig of 1163 bp in length
            *      36234     36333: gap of unknown length
            *      36334     39668: contig of 3335 bp in length
            *      39669     39768: gap of unknown length
            *      39769     42866: contig of 3098 bp in length
            *      42867     42966: gap of unknown length
            *      42967     46088: contig of 3122 bp in length
            *      46089     46188: gap of unknown length
            *      46189     49027: contig of 2839 bp in length
            *      49028     49127: gap of unknown length
            *      49128     51579: contig of 2452 bp in length
            *      51580     51679: gap of unknown length
            *      51680     55192: contig of 3513 bp in length
            *      55193     55292: gap of unknown length
            *      55293     58204: contig of 2912 bp in length
            *      58205     58304: gap of unknown length
            *      58305     60939: contig of 2635 bp in length
            *      60940     61039: gap of unknown length
            *      61040     64184: contig of 3145 bp in length
            *      64185     64284: gap of unknown length
            *      64285     67776: contig of 3492 bp in length
            *      67777     67876: gap of unknown length
            *      67877     72019: contig of 4143 bp in length
            *      72020     72119: gap of unknown length
            *      72120     75147: contig of 3028 bp in length
            *      75148     75247: gap of unknown length
            *      75248     78405: contig of 3158 bp in length
            *      78406     78505: gap of unknown length
            *      78506     82238: contig of 3733 bp in length
            *      82239     82338: gap of unknown length
            *      82339     85153: contig of 2815 bp in length
            *      85154     85253: gap of unknown length
            *      85254     87767: contig of 2514 bp in length
            *      87768     87867: gap of unknown length
            *      87868     90643: contig of 2776 bp in length
            *      90644     90743: gap of unknown length
            *      90744     94013: contig of 3270 bp in length
            *      94014     94113: gap of unknown length
            *      94114     98293: contig of 4180 bp in length
            *      98294     98393: gap of unknown length
            *      98394    101682: contig of 3289 bp in length
            *     101683    101782: gap of unknown length
            *     101783    104237: contig of 2455 bp in length
            *     104238    104337: gap of unknown length
            *     104338    108706: contig of 4369 bp in length
            *     108707    108806: gap of unknown length
            *     108807    113208: contig of 4402 bp in length
            *     113209    113308: gap of unknown length
            *     113309    117238: contig of 3930 bp in length
            *     117239    117338: gap of unknown length
            *     117339    122577: contig of 5239 bp in length
```

```
          *   122578    122677: gap of unknown length
          *   122678    126779: contig of 4102 bp in length
          *   126780    126879: gap of unknown length
          *   126880    131807: contig of 4928 bp in length
          *   131808    131907: gap of unknown length
          *   131908    138315: contig of 6408 bp in length
          *   138316    138415: gap of unknown length
          *   138416    143597: contig of 5182 bp in length
          *   143598    143697: gap of unknown length
          *   143698    152755: contig of 9058 bp in length
          *   152756    152855: gap of unknown length
          *   152856    157897: contig of 5042 bp in length
          *   157898    157997: gap of unknown length
          *   157998    167260: contig of 9263 bp in length
          *   167261    167360: gap of unknown length
          *   167361    178954: contig of 11594 bp in length.
FEATURES             Location/Qualifiers
     source          1..178954

  Query Match             72.1%;  Score 4441.8;  DB 2;  Length 178954;
  Best Local Similarity   99.1%;  Pred. No. 0;
  Matches 4517;  Conservative    0;  Mismatches   34;  Indels    6;  Gaps    5;

Qy        297 GTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCC 356
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174384 GTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCC 174443

Qy        357 AACTCCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATT 416
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174444 AACTCCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATT 174503

Qy        417 GAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAG 476
              |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db     174504 GAACAGATCCCTGCTTCAGAAAGCAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAG 174563

Qy        477 GCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAG 536
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174564 GCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAG 174623

Qy        537 AAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGC 596
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174624 AAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGC 174683

Qy        597 GAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCC 656
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174684 GAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCC 174743

Qy        657 GTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGC 716
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174744 GTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGC 174803

Qy        717 ACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCG 776
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174804 ACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCG 174863

Qy        777 GCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCC 836
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174864 GCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCC 174923

Qy        837 CCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTC 896
              ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     174924 CCAGAATCCCCGAGCAGCCCAGTCAGCAGTGAGACAACATCTTCCTTCAAGAAGTTCTTC 174983

Qy        897 ACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGAT 956
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     174984 ACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGAT 175043

Qy        957 CTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAG 1016
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175044 CTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAG 175103

Qy       1017 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 1076
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175104 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 175163

Qy       1077 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 1136
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175164 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 175223

Qy       1137 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 1196
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175224 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 175283

Qy       1197 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 1256
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175284 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 175343

Qy       1257 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 1316
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175344 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 175403

Qy       1317 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 1376
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175404 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 175463

Qy       1377 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 1436
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175464 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 175523

Qy       1437 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 1496
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175524 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 175583

Qy       1497 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 1556
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175584 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 175643

Qy       1557 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 1616
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175644 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 175703

Qy       1617 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 1676
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175704 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 175763

Qy       1677 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 1736
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175764 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 175823

Qy       1737 GGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAG 1796
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175824 GGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAG 175883

Qy       1797 AAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 1856
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175884 AAGAAGAGGGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 175943

Qy       1857 CGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 1916
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     175944 CGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 176003

Qy       1917 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 1976
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     176004 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 176063

Qy       1977 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 2036
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     176064 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 176123

Qy       2037 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAA 2096
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     176124 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAA 176183

Qy       2097 GGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAA 2156
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     176184 GGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAA 176243
```

```
Qy     2157 GAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGT 2216
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176244 GAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGT 176303

Qy     2217 TCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCA 2276
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176304 TCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCA 176363

Qy     2277 TTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAA 2336
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176364 TTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAA 176423

Qy     2337 GACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTT 2396
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176424 GACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTT 176483

Qy     2397 TCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAA 2456
            |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   176484 TCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAGAACAA 176543

Qy     2457 GCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCC 2516
            |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   176544 GCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTGATGTCCCAGCC 176603

Qy     2517 GTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAAT 2576
            |||||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db   176604 GTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAAT 176663

Qy     2577 ACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTG 2636
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176664 GCGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTG 176723

Qy     2637 GTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAG 2696
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176724 GTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAG 176783

Qy     2697 CGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAA 2756
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176784 CGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAA 176843

Qy     2757 GCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTG 2816
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176844 GCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTG 176903

Qy     2817 CTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTG 2876
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176904 CTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTG 176963

Qy     2877 GATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAA 2936
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   176964 GATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAA 177023

Qy     2937 GTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTG 2996
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177024 GTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTG 177083

Qy     2997 ACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTG 3056
            ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||
Db   177084 ACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGGGTGGTGTG 177143

Qy     3057 CTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAG 3116
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177144 CTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAG 177203

Qy     3117 GTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCA 3176
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177204 GTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCA 177263

Qy     3177 CTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAG 3236
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177264 CTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAG 177323

Qy     3237 GACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCT 3296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177324 GACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCT 177383

Qy     3297 GAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGAC 3356
            ||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   177384 GAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAGCAGATGTAGAC 177443

Qy     3357 CATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCT 3416
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177444 CATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCT 177503

Qy     3417 GAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCA 3476
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177504 GAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCA 177563

Qy     3477 GACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACT 3536
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177564 GACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACT 177623

Qy     3537 GCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAG 3596
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177624 GCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAG 177683

Qy     3597 CCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGA 3656
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177684 CCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGA 177743

Qy     3657 GATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAG 3716
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   177744 GATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCGTTCTGGCAAAG 177803

Qy     3717 ACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAA 3776
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177804 ACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAA 177863

Qy     3777 CAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGAC 3836
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177864 CAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGAC 177923

Qy     3837 AGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTT 3896
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177924 AGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTT 177983

Qy     3897 AGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCA 3956
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   177984 AGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCA 178043

Qy     3957 GAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTAC 4016
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178044 GAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTAC 178103

Qy     4017 GGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGC 4076
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178104 GGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGC 178163

Qy     4077 CTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTT 4136
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178164 CTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTT 178223

Qy     4137 CAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACG 4196
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178224 CAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACG 178283

Qy     4197 GTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCT 4256
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178284 GTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCT 178343

Qy     4257 GAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCC 4316
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178344 GAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCC 178403

Qy     4317 CTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGC 4376
```

```
Db 178404 CTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGC 178463

Qy   4377 ACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAA 4436
Db 178464 ACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAA 178523

Qy   4437 GAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAA 4496
Db 178524 GAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAA 178583

Qy   4497 GTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTG 4556
Db 178584 GTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTG 178643

Qy   4557 AACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCAT 4616
Db 178644 AACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCAT 178703

Qy   4617 GCTTATGATTCACAGACCCAGGTTCCTGC-ATGCAGGCTTGACAGCAGGGAGCCCAACAG 4675
Db 178704 GCTTATGATTCACAGACCCAGGTTCCTGCAATGCAGGCTGACAGCCAGGGAGCCCAACAG 178763

Qy   4676 ATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGA 4735
Db 178764 ATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGA 178823

Qy   4736 CTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTT 4795
Db 178824 CTTGCAAGTCCTGACCGTTCTGGA-GCATGGGCTCAGCCTCCGAATGCTTGGCCGCGC-T 178881

Qy   4796 GCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAA 4852
Db 178882 GCANNTGAAAGCGCC-GTGTCAAAGTA--GCATGGAGAGCTGCCTCCTCACCCAAGA 178935


RESULT 7
AF326230
LOCUS       AF326230               5902 bp    mRNA    linear   ROD 16-JUL-2001
DEFINITION  Mus musculus testis-specific SSeCKS mRNA, complete cds.
ACCESSION   AF326230
VERSION     AF326230.1  GI:13236407
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 5902)
  AUTHORS   Camus,A., Mesbah,K., Rallu,M., Babinet,C. and Barra,J.
  TITLE     Gene trap insertion reveals two open reading frames in the mouse
            SSeCKS gene: the form predominantly detected in the nervous system
            is suppressed by the insertion while the other, specific of the
            testis, remains expressed
  JOURNAL   Mech. Dev. 105 (1-2), 79-91 (2001)
  MEDLINE   21322683
   PUBMED   11429284
REFERENCE   2  (bases 1 to 5902)
  AUTHORS   Mesbah,K., Babinet,C. and Barra,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2000) Immunologie, Institut Pasteur, 25 rue du
            Docteur Roux, Paris 75015, France
FEATURES             Location/Qualifiers
     source          1..5902
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /tissue_type="testis"
     CDS             160..4899
                     /codon_start=1
                     /product="testis-specific SSeCKS"
                     /protein_id="AAK16152.1"
                     /db_xref="GI:13236408"
                     /translation="MAASSTVVEDITKDEQEETPEIIEQIPASESNVEEMAQAAESQA
                     NDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPGVET
                     VGESASKESELKQSTEKQEGTLKQAQSSTEIPLQAESGQGTEEEAAKDGEENREKEPT
                     KPLESPTSPVSNETTSSFKKFFTHGWAGWRKKTSFKKPKEDDLETSEKRKEQEAEKVD
                     EEEGEKTEPAPAEEQEPAEGTDQARLSADYEKVELPLEDQVGDLEALSEKCAPLATEV
                     FDEKTEAHQEVVAEVHVSTVEKMTKGQGGAEVEGDVVVEGSGESLPPEKLAETQEVPQ
                     EAEPVEELMKTKEVCVSGGDHTQLTDLSPEEKMLPKHPEGIVSEVEMLSSQERIKVQG
                     SPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIQTESPESADEQKGESSASS
                     PEEPEEIACLEKGPSEAPQEAEAEEGATSDGEKKREGITPWASFKKMVTPKKRVRRPS
                     ESDKEEELDKVKSATLSSTESTASGMQDEVRAVGEEQRSEEPKRRVDTSVSWEALICV
                     GSSKKRARKASSSDDEGGPRTLGGDGHRAEEASKDKEADALPASTQEQDQAHGSSSPE
                     PAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEERAEDSGAEQLASEIEPSREESWVSI
                     KKFIPGRRKKRADGKQEQAAVEDSGPGEINEDDPDVPAVVPLSEYDAVEREKLEAQRA
                     QENVELPQLKGAVYVSEELSKTLVHTVSVAVIDGTRAVTSAEERSPSWISASMTEPLE
                     HAEGVATPPVGEVTEKDITAEATPALAQTLPGGKDAHDDIVTSEVDFTSEAVTAAETT
                     EALRAEELTEASGAEETTDMVSAVSQLSDSPDTTEEATPVQEVEGGMLDTEEQERQTQ
                     AVLQAVADKVKEDSQVPATQTLQRAGPKALEKVEEVEEDSEVLATEKEKDVVPEGPVQ
                     EAETEHLAQGSETVQATPESLEVPEVTEDVDRATTCQVIKHQQLMEQAVAPESSETLT
                     DSETNGSTPLADSDTPNGTQQDETVDSQDSNAIAAVKQSQVTEEEAAAAQTEGPSTPS
                     SFPAQEEHREKPGRDVLEPTQALAAGAVPILAKAEVGQEGEAGQFDGEKVKDGQCVKE
                     LEVPVHTGPNSQKTADLTRDSEVMEVARCQETESNEEQSISPEKREMGTDVEKEETET
                     KTEQASEEHEQETAAPEHEGTHPKPVLTADMPHSERGKALGSLEGSPSLPDQDKADCI
                     EVQVQSSDTPVTQTTEAVKKVEETVATSEMDESLECAGAQSLPAEKLSETGGYGTLQH
                     GEDTVPQGPESQAESIPIIVTPAPESILHSDLQREVSASQKQRSDEDNKPDAGPDAAG
                     KESAAREKILRAEPEILELESKSNKIVQSVIQTAVDQFARTETAPETHASDLQNQVPV
                     MQADSQGAQQMLDKDESDLQVSPQDGTLSAVAQEGLAVSDSSEGMSKASEMITTLAVE
                     SASVKESVEKLPLQCKDEKEHAADGPQHQSLAKAEADASGNLTKESPDTNGPKLTEEG
                     DALKEEMNKAQTEEDDLQEPKGDLTES"
BASE COUNT     1774 a   1351 c   1677 g   1100 t
ORIGIN

  Query Match             69.9%;  Score 4304;  DB 10;  Length 5902;
  Best Local Similarity   86.3%;  Pred. No. 0;
  Matches 5105;  Conservative    0;  Mismatches  620;  Indels  189;  Gaps   24;

Qy        298 TTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCA 357
Db        110 TTGGACAGAGAGAGTCAGAAGATGTGAAAGAAAAAGACCGAGCTAAAGAAATGGCAGCCA 169

Qy        358 ACTCCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTG 417
Db        170 GTTCCACAGTTGTTGAAGATATCACAAAGGACGAGCAGGAGGAAACACCGGAAATAATCG 229

Qy        418 AACAGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGG 477
Db        230 AACAGATCCCTGCTTCAGAGAGCAATGTGGAAGAAATGGCGCAGGCTGCTGAGTCCCAAG 289

Qy        478 CTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGA 537
Db        290 CTAATGACGTCGGCTTCAAGAAGGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGA 349

Qy        538 AGGATAAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCG 597
Db        350 AGGATAAAAACGAAAAGTCAGATACCGTCCAGCTACTCACTGTCAAGAAGGATGAAGGCG 409

Qy        598 AAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCG 657
Db        410 AAGGGGCAGAAGCCTCCGTCGGAGCAGGAGACCACCAAGAGCCCGGAGTGGAGA---CCG 466

Qy        658 TCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCA 717
Db        467 TCGGCGAATCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCA 526

Qy        718 CCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGG 777
Db        527 CCCTGAAGCAAGCACAGAGCAGCACAGAAATTCCCCTTCAAGCCGAATCTGGTCAAGGGA 586

Qy        778 CTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCC 837
Db        587 CCGAGGAAGAAGCAGCCAAAGATGGAGAAGAAAACCGAGAGAAAGAACCTACCAAGCCCC 646

Qy        838 CAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCA 897
Db        647 TAGAATCTCCGACCAGCCCTGTCAGCAATGAGACAACATCTTCCTTCAAGAAATTCTTCA 706
```

```
Qy    898 CTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATC 957
Db    707 CTCACGGGTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAACCAAAGGAAGATGATC 766

Qy    958 TGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGG 1017
Db    767 TGGAAACTTCCGAGAAGAGAAAGGAGCAAGAGGCTGAAAAAGTAGACGAGGAAGAAGGGG 826

Qy   1018 AAAAGACAGAGCCAGCC---TCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCA 1074
Db    827 AAAAGACAGAGCCAGCCCCAGCCGAGGAGCAGGAGCCTGCAGAAGGCACAGACCAGGCCA 886

Qy   1075 GGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGG 1134
Db    887 GGTTGTCAGCCGACTATGAGAAGGTGGAGTTGCCTTTGGAAGACCAGGTCGGTGACCTGG 946

Qy   1135 AGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGG 1194
Db    947 AGGCATTGTC---GGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGACGG 1003

Qy   1195 AAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGG 1254
Db   1004 AAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGATGACGAAAG 1063

Qy   1255 AGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCT 1314
Db   1064 GGCA---AGGAGGAGCAGAGGTGGAAGGGGATGTGGTGGTGGAAGGATCGGGAGAATCCT 1120

Qy   1315 TGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGG 1374
Db   1121 TGCCCCCTGAGAAACTGGCTGAGACCCAGGAGGTCCCCCAGGAAGCTGAGCCTGTGGAGG 1180

Qy   1375 AGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACC 1434
Db   1181 AGCTGATGAAGACCAAAGAAGTATGCGTCTCTGGGGGTGACCATACTCAGCTGACAGATC 1240

Qy   1435 TAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGA 1494
Db   1241 TAAGTCCTGAAGAGAAGATGCTACCCAAACACCCCGAAGGCATTGTCAGTGAGGTGGAGA 1300

Qy   1495 TGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTA 1554
Db   1301 TGCTGTCCTCTCAGGAGAGAATCAAGGTACAGGGAAGTCCCCTGAAGAAGCTCTTCAGCA 1360

Qy   1555 GCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAA---ACGAGGAGGTGGGG 1611
Db   1361 GTTCGGGCTTAAAGAAGCTCTCCGGGAAGAAGCAGAAGGGGAAGAGAGGAGGAGGCGGGG 1420

Qy   1612 GAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATG 1671
Db   1421 GAGATGAAGAGCCAGGAGAATACCAACACATTCAAACCGAGTCCCCAGAGAGTGCTGACG 1480

Qy   1672 AGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGG 1731
Db   1481 AGCAGAAGGGAGAGAGCTCTGCCTCTTCCCCTGAAGAGCCCGAGGAGATCGCGTGTCTGG 1540

Qy   1732 AGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATG 1791
Db   1541 AGAAGGGGCCATCGGAAGCACCCCAGGAAGCGGAAGCTGAGGAAGGAGCGACTTCCGACG 1600

Qy   1792 GAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCA 1851
Db   1601 GAGAGAAGAAAAGGAAGGGATCACCCCCTGGGCATCCTTCAAAAAGATGGTGACACCCA 1660

Qy   1852 AGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGA 1911
Db   1661 AGAAACGGGTCCGAAGACCTTCTGAGAGCGACAAGGAAGAAGAGCTGGATAAGGTCAAGA 1720

Qy   1912 GCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTG 1971
Db   1721 GTGCCACCTTGTCCTCCACGGAGAGCACGGCGTCTGGAATGCAGGATGAGGTCAGAGCGG 1780

Qy   1972 TTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGG 2031
Db   1781 TTGGCGAGGAGCAAAGGTCAGAGGAGCCAAAGCGCAGGGTGGATACTTCAGTGTCTTGGG 1840

Qy   2032 AAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATG 2091
Db   1841 AGGCGTTGATTTGTGTCGGATCGTCCAAGAAGAGAGCGAGGAAGGCATCCTCTTCAGATG 1900

Qy   2092 ATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAG 2151
Db   1901 ATGAAGGAGGGCCAAGAACACTGGGAGGGGATGGCCACAGAGCGGAGGAGGCTAGCAAAG 1960

Qy   2152 ACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAG 2211
Db   1961 ACAAAGAAG------CAGATGCTCTTCCTGCCAGCACCCAGGAACAAGACCAAGCGCACG 2014

Qy   2212 GAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGG 2271
Db   2015 GAAGTTCCTCACCCGAGCCAGCTGGAAGCCCTTCTGAAGGGGAGGGCGTCTCCACCTGGG 2074

Qy   2272 AGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAG 2331
Db   2075 AGTCATTTAAGAGATTAGTCACTCCACGAAAAAAATCCAAGTCAAAACTGGAAGAGAGAG 2134

Qy   2332 CCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTT 2391
Db   2135 CCGAAGACTCCGGTGCAGAGCAGTTGGCCTCCGAGATCGAACCAAGTAGAGAGGAATCTT 2194

Qy   2392 GGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAG 2451
Db   2195 GGGTTTCCATTAAGAAATTTATTCCTGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAG 2254

Qy   2452 AACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCC 2511
Db   2255 AACAGGCCGCCGTTGAAGACTCGGGGCCAGGAGAGATCAATGAGGACGACCCCGACGTCC 2314

Qy   2512 CAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGA-------- 2563
Db   2315 CAGCTGTTGTGCCTCTGTCTGAGTACGATGCGGTAGAGAGAGAGAAGCTGGAAGCGCAGC 2374

Qy   2564 -AGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGC 2622
Db   2375 GAGCTCAGGAGAACGTGGAGCTGCCCCAGCTGAAGGGGGCTGTGTATGTGTCTGAGGAGC 2434

Qy   2623 TCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCA 2682
Db   2435 TTAGTAAGACTCTGGTTCACACTGTGAGTGTCGCGGTCATTGATGGGACCAGGGCAGTCA 2494

Qy   2683 CCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAAC 2742
Db   2495 CCAGTGCCGAAGAGCGGTCCCCTTCGTGGATATCTGCTTCCATGACAGAACCTCTTGAGC 2554

Qy   2743 ACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAG 2802
Db   2555 ACGCAGAGGGAGTGGCCACACCGCCTGTTGGAGAGGTCACTGAAAAAGACATCACTGCAG 2614

Qy   2803 AAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGG 2862
Db   2615 AAGCAACTCCTGCACTCGCCCAGACTTTACCAGGGGGCAAAGATGCCCATGACGACATAG 2674

Qy   2863 TCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTC 2922
Db   2675 TCACCAGTGAAGTGGATTTTACCTCAGAAGCAGTGACAGCCGCAGAAACCACAGAGGCGC 2734

Qy   2923 TCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCG 2982
Db   2735 TCCGCGCTGAAGAACTTACCGAAGCATCAGGGGCAGAAGAGACCACAGACATGGTGTCTG 2794

Qy   2983 CAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGG 3042
Db   2795 CAGTTTCCCAGCTGTCCGACTCCCCGGACACCACAGAGGAAGCCACCCCAGTTCAGGAGG 2854

Qy   3043 TAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAG 3102
```

```
Db   2855 TAGAGGGTGGCATGCTAGATACGGAAGAACAGGAGCGCCAGACGCAGGCCGTCCTCCAAG 2914

Qy   3103 CCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAA 3162
Db   2915 CCGTTGCAGACAAAGTGAAAGAGGACTCCCAGGTGCCTGCAACCCAGACTCTGCAGAGAG 2974

Qy   3163 CGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTT 3222
Db   2975 CAGGGCCGAAAGCACTGGAGAAGGTGGAGGAGGTAGAGGAGGACTCCGAGGTGCTGGCTA 3034

Qy   3223 CGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATC 3282
Db   3035 CCGAGAAAGAGAAGGATGTTGTGCCGGAAGGACCCGTGCAGGAAGCTGAAACTGAGCATC 3094

Qy   3283 TTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCA 3342
Db   3095 TTGCACAGGGCTCCGAGACTGTACAGGCTACCCCAGAGAGCCTTGAAGTTCCTGAAGTCA 3154

Qy   3343 CGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAAC 3402
Db   3155 CAGAGGATGTAGACCGTGCCACCACATGCCAGGTTATCAAGCACCAGCAGCTGATGGAAC 3214

Qy   3403 AGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTC 3462
Db   3215 AGGCTGTGGCCCCTGAGTCATCTGAAACCTTGACAGACAGTGAGACAAATGGAAGTACTC 3274

Qy   3463 CCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGG 3522
Db   3275 CCCTCGCAGATTCAGACACTCCAAACGGGACACAGCAAGACGAGACCGTTGACAGCCAGG 3334

Qy   3523 ACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTG 3582
Db   3335 ACAGTAATGCCATTGCCGCCGTCAAGCAGTCACAGGTCACTGAAGAGGAGGCAGCTGCTG 3394

Qy   3583 CTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGG 3642
Db   3395 CTCAGACGGAGGGGCCTTCAACACCATC---TAGTTTTCCAGCCCAGGAAGAACACAGGG 3451

Qy   3643 AAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGC 3702
Db   3452 AAAAACCAGGAAGGGATGTTCTAGAACCCACA---CAAGCGCTGGCTGCCGGGGCAGTGC 3508

Qy   3703 CCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAA 3762
Db   3509 CTATTCTGGCAAAGGCTGAGGTGGGTCAAGAGGGTGAGGCTGGCCAGTTTGATGGAGAAA 3568

Qy   3763 AAGTCAAAG---------------AAGAACAGGAGGTGTTTGTACACTCTGGACCCAACA 3807
Db   3569 AAGTCAAAGACGGACAGTGTGTTAAAGAACTGGAGGTGCCTGTGCACACTGGACCCAACA 3628

Qy   3808 GTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGG 3867
Db   3629 GTCAAAAGACTGCTGACTTGACACGTGACAGTGAAGTAATGGAAGTGGCCAGATGTCAGG 3688

Qy   3868 AAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACG 3927
Db   3689 AAACTGAGAGTAATGAAGAACAGAGTATTAGCCCGGAGAAAAGAGAGATGGGAACCGACG 3748

Qy   3928 TTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAA 3987
Db   3749 TTGAAAAGGAGGAAACAGAGACCAAGACAGAGCAAGCCAGTGAAGAACATGAGCAGGAAA 3808

Qy   3988 CAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCA 4047
Db   3809 CAGCTGCTCCTGAGCATGAAGGAACCCACCCTAAGCCAGTCCTGACAGCTGACATGCCTC 3868

Qy   4048 GCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAG 4107
Db   3869 ACTCAGAGAGGGGAAAGGCACTGGGCAGCCTTGAAGGAAGCCCTTCTCTCCCAGACCAAG 3928

Qy   4108 ACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAG 4167

Db   3929 ACAAAGCAGATTGCATAGAGGTTCAAGTTCAAAGCTCAGACACACCAGTCACTCAAACAA 3988

Qy   4168 CAGAAGCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAG 4227
Db   3989 CCGAAGCTGTGAAAAAGGTCGAAGAAACTGTGGCAACTTCAGAGATGGATGAAAGTTTGG 4048

Qy   4228 AGTGTGTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGA 4287
Db   4049 AGTGTGCAGGTGCGCAATCATTACCAGCTGAGAAGCTCTCCGAAACCGGTGGCTACGGGA 4108

Qy   4288 CTCTTCAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCC 4347
Db   4109 CTCTTCAGCATGGAGAGGACACCGTGCCCCAGGGGCCTGAGTCTCAGGCAGAGTCCATCC 4168

Qy   4348 CAATCATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAA 4407
Db   4169 CAATAATAGTAACTCCTGCTCCTGAAAGCATCCTACATTCTGACCTTCAAAGAGAAGTGA 4228

Qy   4408 GCGCATCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTG 4467
Db   4229 GCGCATCCCAGAAACAGAGATCAGATGAAGATAACAAGCCAGATGCTGGTCCTGATGCTG 4288

Qy   4468 ACGGCAAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAAC 4527
Db   4289 CCGGCAAGGAGAGTGCAGCAAGAGAGAAAATCCTCAGGGCTGAACCTGAGATCTTGGAAC 4348

Qy   4528 TTGAGAGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCG 4587
Db   4349 TTGAGAGTAAGAGCAATAAGATTGTCCAGAGTGTCATCCAGACAGCCGTCGACCAGTTTG 4408

Qy   4588 CACGTACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTG-CA 4646
Db   4409 CACGTACAGAAACAGCCCCCGAAACCCACGCTTCTGATTTACAGAATCAGGTTCCTGTGA 4468

Qy   4647 TGCAGGCTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATG-------- 4698
Db   4469 TGCAGGCTGACAGCCAGGGAGCACAACAGATGCTGGACAAAGATGAAAGCGACCTTCAAG 4528

Qy   4699 -------CCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACC 4751
Db   4529 TCTCCCCCCAAGATGGAACACTCAGTGCCGTAGCCCAGGAAGGACTTGCGGTTTCTGATA 4588

Qy   4752 GTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCG 4811
Db   4589 GTTCTGAAGGCATGAGCAAGGCTTCAGAAATGATCACCACGCTTGCAGTTGAAAGTGCCA 4648

Qy   4812 GTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCT 4871
Db   4649 GTGTCAAAGAAAGTGTTGAGAAGCTGCCTCTTCAGTGCAAAGAT-GAAAAGGAACATGCT 4707

Qy   4872 GCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTA 4931
Db   4708 GCTGACGGCCCCCAGCACCAAAGCTTAGCCAAGGCAGAGGCGGATGCCTCTGGAAATCTA 4767

Qy   4932 ACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAA 4991
Db   4768 ACCAAAGAGTCCCCAGACACCAACGGACCAAAGCTAACCGAGGAAGGAGATGCCCTGAAA 4827

Qy   4992 GTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCA 5051
Db   4828 GAAG-----------------------------------AAATGAACAAGGCCCAGACA 4851

Qy   5052 GAAGAG---GACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGC 5108
Db   4852 GAAGAGGACGACCTACAGGAGCCAAAGGGAGACCTGACAGAATCCTAAGACGTGAGTTGC 4911

Qy   5109 TCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTT 5168
Db   4912 TCAT--------TGTAAAACTAGAATGTGAAGTGAAGTCACAGAACAAGATGCTGCTGTT 4963

Qy   5169 GGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGA 5228
Db   4964 GGGACCTTGAGACCAAAATTTCAGAGCCCTTGAGGTGCAGAGAGCAGAGCCGTCCAATGA 5023
```

```
Qy    5229 TTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTA 5288
Db    5024 TTTCAACCCTGCAGAGCACCCCGACAATCCTGAGGCTGCATCGGGAGCTAGAGCCAGCTA 5083

Qy    5289 ACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTA 5348
Db    5084 ACATTTCCTCTTTTCAAGACTGCCCTTGATTTGCCCCTTGGTG---------------- 5126

Qy    5349 ACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGG 5408
Db    5127 ---------------------------------------------CCATGTATTTCAG 5139

Qy    5409 ATTTAAGGTCCTGCGTTCTCAACCTGGAACCAAT-TCTGCCATACCTAGTTCCACTTCTC 5467
Db    5140 AGTTAAGGTCCTGCTTTCTCCACCTGGAACCAATGTCGGCAATACCTAGTCCCACTTCTC 5199

Qy    5468 AAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTA 5527
Db    5200 AAACTGGAGCCTCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTA 5259

Qy    5528 CCTATTGTATA-TTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATA 5586
Db    5260 CCTATTGTATATTTTTTTTCTAAGGTTTAAGCACATGCCTTTTGTATTATGCAGTATATA 5319

Qy    5587 ACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGC 5646
Db    5320 ATGGGTGTGCAGCCATAGCGAAGCTTTGAGAAGCCCCAAGCCTCAACTGTAACCTGCAGC 5379

Qy    5647 AAACAG----ATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGC 5702
Db    5380 AAACAGAGAAATAACATTCCTGGCAGGAAGATACAAGTC-TTTTAAAGTTTACTGATGC 5438

Qy    5703 TTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCA 5762
Db    5439 TTAACTCTGTGGGCTCGTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGTGAGCTCA 5498

Qy    5763 GAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCC 5822
Db    5499 GAAATAAAAACTCCATTTTGAGACATTCAGAATGTCCCAACATTACGACAACATTTTTTT 5558

Qy    5823 CCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCT 5882
Db    5559 TTTTTTTTTCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTTAGTGTCAGATTAAGCCCC 5618

Qy    5883 GTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAG 5942
Db    5619 ATCTCTTAACAATATGGACAGATGAGTGTGCC-ATGGCCATGAGCTATTTCCTAACGCAG 5677

Qy    5943 AAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGC 6002
Db    5678 AAGGAATTTGTTGT-TACTTTTTGGATTGTACTCTTCTAGGCTGGACCGAATTCATATGC 5736

Qy    6003 AGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTG 6062
Db    5737 AGATTGAAGTGAGACCTGCTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTG 5796

Qy    6063 TGTTATATCTGTGCCCCTTC-TTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATT 6121
Db    5797 TATTATATCTGTGCCCCTTCTTTTAAGAACAATGTTACATTATGTTCCTTTGGATAAATT 5856

Qy    6122 GTGATTTGACAACTGATTTAAATAAACATATTTG 6155
Db    5857 GTGATTTGACAACTGATTTCAATAAAAATATTTG 5890
```

```
RESULT 8
AX405857
LOCUS       AX405857                6614 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 272 from Patent WO0222660.
ACCESSION   AX405857
VERSION     AX405857.1  GI:21439153
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F.,
            Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.
  TITLE     Novel nucleic acids and polypeptides
  JOURNAL   Patent: WO 0222660-A 272 21-MAR-2002;
            HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1..6614
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             192..5543
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD34964.1"
                     /db_xref="GI:21439154"
                     /translation="MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADP
                     AIAASDPATKLLQKNGQLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEV
                     IVTEVGQRDSEDVSERDSDKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELT
                     QPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVKKDEGEGAAGAGDHQDP
                     SLGAGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAVEECKEEGEEKQEKE
                     PSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEK
                     VDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPS
                     EEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTEVEETAGSVPAEELVEM
                     DAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQER
                     MKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGGDEESGEHTQVPADSPDSQEEQKGESS
                     ASSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVR
                     RPSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEAL
                     ICVGSSKKRARRRSSSDEEGGPKAMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQG
                     SSSPEQAGSPTEGEGVSTWESFKRLVTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEP
                     GKEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVE
                     REKMEAQQAQKGAEQPEQKAATEVSKELSESQVHMMAAAVADGTRAATIIEERSPSWI
                     SASVTEPLEQVEAEAALLTEEVLEREVIAEEEPPTVTEPLPENREARGDTVVSEAELT
                     PEAVTAAETAGPLGSEEGTEASAAEETTEMVSAVSQLTDSPDTTEEATPVQEVEGGVP
                     DIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDVLQPVQRAEAERPEEQAEASGLKK
                     ETDVVLKVDAQEAKTEPFTQGKVVGQTTPESFEKAPQVTESIESSELVTTCQAETLAG
                     VKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIHEENEVASGT
                     QSGGTEAEAVPAQKERPPAPSSFVFQEETKEQSKMEDTLEHTDKEVSVETVSILSKTE
                     GTQEADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDAL
                     ELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSKQLLQTV
                     NVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLGETAN
                     ILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATTKK
                     GLSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEEDLEPENGILELETKSSKLVQ
                     NIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQETEKEGEDPQASAQDET
                     PITSAKEESESTAVGQAHSDISKDMSEASEKTMTVEVEGSTVNDQQLEEVVLPSEEEG
                     GGAGTKSVPEDDGHALLAERIEKSLVEPKEDEKGDDVDDPENQNSALADTDASGGLTK
                     ESPDTNGPKQKEKEDAQEVELQEGKVHSESDKAITPQAQEELQKQERESAKSELTES"
BASE COUNT     2094 a   1420 c   1827 g   1273 t
ORIGIN

  Query Match             35.4%;  Score 2179.6;  DB 6;  Length 6614;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 4223;  Conservative    0;  Mismatches 1794;  Indels  537;  Gaps   43;
```

```
Qy       3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db     162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy      63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db     222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy     120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db     282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy     171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db     342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401
```

```
Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy    269 -------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACAGAGGTTGGACAGAGA 521

Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    582 GTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCT 641

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    642 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 701

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    702 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 761

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    762 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 821

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    822 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 866

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    867 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 926

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    927 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCA---GTGGAGGAA 983

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    984 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1043

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1044 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1103

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1104 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1163

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1164 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1223

Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1224 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1283

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1284 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1343

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1344 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1403

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1404 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1463

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1464 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1499

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1500 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAA 1559

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1560 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1619

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1620 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1679

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1680 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1739

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1740 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1799

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1800 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1856

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1857 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1916

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1917 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1976

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1977 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2036

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2037 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2096

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2097 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2156

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2157 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2216

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2217 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2276

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2277 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2336

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2337 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2396

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2397 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2456

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2457 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2516

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
```

```
Db  2517 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2576

Qy  2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db  2577 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2636

Qy  2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db  2637 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2696

Qy  2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db  2697 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2756

Qy  2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db  2757 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2816

Qy  2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db  2817 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2876

Qy  2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db  2877 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2936

Qy  2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db  2937 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2996

Qy  2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db  2997 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3056

Qy  2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db  3057 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3116

Qy  2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db  3117 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3176

Qy  2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db  3177 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3236

Qy  3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db  3237 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3296

Qy  3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db  3297 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3356

Qy  3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db  3357 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3416

Qy  3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db  3417 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3476

Qy  3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db  3477 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3536

Qy  3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db  3537 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3596

Qy  3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db  3597 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3656

Qy  3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db  3657 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3716

Qy  3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db  3717 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3776

Qy  3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db  3777 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3836

Qy  3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db  3837 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3896

Qy  3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db  3897 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3956

Qy  3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db  3957 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4016

Qy  3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db  4017 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4076

Qy  3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db  4077 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4136

Qy  3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db  4137 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4196

Qy  3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db  4197 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4256

Qy  3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db  4257 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4316

Qy  4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db  4317 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4376

Qy  4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db  4377 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4436

Qy  4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db  4437 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4496

Qy  4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db  4497 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4556

Qy  4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db  4557 AAATCCTCTGAAAAAAATGAAGACTTTGCCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4616

Qy  4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db  4617 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4676

Qy  4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db  4677 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4736
```

```
Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4737 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGGAGGAT 4790

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4791 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4850

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4851 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4910

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4911 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4970

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4971 GAAACGGAGAAAGAAGGAGAGGACCCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5030

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5031 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5090

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5091 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5150

Qy   4819 ------------------------------------------------------------ 4818
Db   5151 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5210

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5211 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5270

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5271 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5330

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5331 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5390

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5391 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5450

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5451 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5510

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5511 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5570

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5571 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5629

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5630 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5688

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5689 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5747

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5748 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5787

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5788 ------------------------------------------CATATAAATTCTGATTTA 5805

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5806 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5865

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5866 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5922

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5923 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5982

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5983 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6042

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6043 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6102

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6103 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6162

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6163 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6222

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6223 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6275

Qy   5867 TGTCAGA----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6276 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6335

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6336 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6395

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6396 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6453

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6454 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6513

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6514 TGCCATTCTTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6573

Qy   6134 CTGATTTAAATAAA 6147
Db   6574 CTGATTTAAATAAA 6587
```

RESULT 9
AX329766

```
LOCUS       AX329766                 6608 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 275 from Patent WO0194629.
ACCESSION   AX329766
VERSION     AX329766.1  GI:18102744
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
```

```
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 275 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1..6608
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     2091 a   1418 c   1827 g   1272 t
ORIGIN

  Query Match             34.8%;  Score 2144.4;  DB 6;  Length 6608;
  Best Local Similarity   64.3%;  Pred. No. 0;
  Matches 4215;  Conservative    0;  Mismatches 1796;  Indels  543;  Gaps   45;

Qy          3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db        162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy         63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db        222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy        120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db        282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy        171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db        342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy        231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db        402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy        269 -------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db        462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy        309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db        522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy        369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db        582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy        429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db        639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy        489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db        699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy        549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db        759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy        609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db        819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy        669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db        864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy        729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db        924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy        789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db        981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy        849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db       1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy        909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db       1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy        969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db       1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy       1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db       1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy       1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db       1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy       1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db       1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy       1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db       1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy       1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db       1461 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1496

Qy       1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db       1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy       1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db       1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy       1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db       1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy       1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db       1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy       1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db       1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy       1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db       1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy       1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db       1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy       1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db       1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy       1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db       1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy       1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
```

```
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3833

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3834 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3893

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3894 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3953

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3954 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4013

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4014 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4073

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4074 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4133

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db   4134 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4193

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4194 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4253
```

```
Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4254 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4313

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4314 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4373

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4374 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4433

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4434 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4493

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4494 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4553

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4554 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4613

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4614 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4673

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4674 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4733

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4734 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4784

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4785 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4844

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4845 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4904

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4905 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4964

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4965 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5024

Qy   4721 GCAGCCCA---------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5025 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5084

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5085 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5144

Qy   4819 ------------------------------------------------------------ 4818
Db   5145 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5204

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5205 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5264

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5265 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5324

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5325 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5384

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5385 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5444

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5445 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5504

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5505 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5564

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5565 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5623

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5624 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5682

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5683 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5741

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5742 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5781

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5782 -----------------------------------------CATATAAATTCTGATTTA 5799

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5800 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5859

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5860 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5916

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5917 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5976

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5977 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6036

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6037 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6096

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6097 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6156

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6157 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6216

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6217 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6269

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6270 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6329

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
```

```
Db   6330 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6389

Qy   5954 TGTACGTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6390 TGTAAG--TTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6447

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6448 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6507

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6508 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6567

Qy   6134 CTGATTTAAATAAA 6147
Db   6568 CTGATTTAAATAAA 6581
```

```
RESULT 10
AX410793
LOCUS       AX410793                6608 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 3440 from Patent WO0229103.
ACCESSION   AX410793
VERSION     AX410793.1  GI:21443498
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 3440 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1..6608
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. U81607"
BASE COUNT     2091 a   1418 c   1827 g   1272 t
ORIGIN

  Query Match             34.8%;  Score 2144.4;  DB 6;  Length 6608;
  Best Local Similarity   64.3%;  Pred. No. 0;
  Matches 4215;  Conservative    0;  Mismatches 1796;  Indels  543;  Gaps   45;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy    269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy   1029 CCAGCCTCGGAGGA-------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1461 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1496

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616
```

```
Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy   1599 CGAGGAGGTGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
```

```
Db  3774 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3833

Qy  3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db  3834 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3893

Qy  3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db  3894 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3953

Qy  3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db  3954 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4013

Qy  3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db  4014 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4073

Qy  3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db  4074 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4133

Qy  3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db  4134 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4193

Qy  3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db  4194 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4253

Qy  3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db  4254 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4313

Qy  4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db  4314 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4373

Qy  4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db  4374 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4433

Qy  4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db  4434 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4493

Qy  4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db  4494 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4553

Qy  4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db  4554 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4613

Qy  4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db  4614 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4673

Qy  4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db  4674 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4733

Qy  4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db  4734 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4784

Qy  4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db  4785 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4844

Qy  4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db  4845 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4904

Qy  4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db  4905 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4964

Qy  4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db  4965 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5024

Qy  4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db  5025 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5084

Qy  4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db  5085 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5144

Qy  4819 ------------------------------------------------------------ 4818
Db  5145 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5204

Qy  4819 ------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db  5205 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5264

Qy  4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db  5265 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5324

Qy  4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db  5325 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5384

Qy  4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db  5385 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5444

Qy  5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db  5445 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5504

Qy  5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db  5505 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5564

Qy  5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db  5565 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5623

Qy  5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db  5624 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5682

Qy  5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db  5683 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5741

Qy  5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db  5742 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5781

Qy  5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db  5782 -----------------------------------------CATATAAATTCTGATTTA 5799

Qy  5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db  5800 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5859

Qy  5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db  5860 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5916
```

```
Qy   5533 TGTATATTTTTTCTAACGTTT-----------------AAGCACATGCTTTTGTATTA 5575
          ||||||||||||||  |   ||                   |  ||||||||||||||  |
Db   5917 TGTATATTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTGTATCA 5976

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
            || |||||  | |||       ||||||  |   |   |||   ||||  ||||||||  |
Db   5977 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6036

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
          ||      ||  |||  |  |  | |  |          |    |   ||   |     || ||
Db   6037 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6096

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
          |||| ||  |||  ||| ||               |   |  || ||||||||  ||||    |||||
Db   6097 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6156

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
          ||      |       |     |||    |||    ||||||||||||  |||  ||||  | |  |
Db   6157 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6216

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
          |     ||   ||||          || ||||  |||||   |||||  ||||||||||      |
Db   6217 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6269

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
             ||   |                       |     |   |   |   |   |||||  || |||
Db   6270 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6329

Qy   5910 GTGCCTAAGGCCATGAGA---------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
          |||| |||   |    |   |                   |   |    | ||| ||||    |   |||
Db   6330 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6389

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
          |||| |   ||||||||| ||||||| |||||||||| | |||||  |       |     ||| |
Db   6390 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6447

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
          ||||  |||||||||| |||||||||||||||||||||||||||| ||| ||||  |  |||| ||
Db   6448 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6507

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
          ||||   | |||||||||||||||||||   |||  |||||||||  ||||||||||||  |
Db   6508 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6567

Qy   6134 CTGATTTAAATAAA 6147
          ||||||||||||||
Db   6568 CTGATTTAAATAAA 6581
```

```
RESULT 11
HSU81607
LOCUS       HSU81607                6608 bp    mRNA    linear   PRI 26-JUN-1997
DEFINITION  Homo sapiens gravin mRNA, complete cds.
ACCESSION   U81607
VERSION     U81607.1  GI:2218076
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6608)
  AUTHORS   Nauert,J.B., Klauck,T.M., Langeberg,L.K. and Scott,J.D.
  TITLE     Gravin, an autoantigen recognized by serum from myasthenia gravis
            patients, is a kinase scaffold protein
  JOURNAL   Curr. Biol. 7 (1), 52-62 (1997)
  MEDLINE   97153077
   PUBMED   9000000
REFERENCE   2  (bases 1 to 6608)
  AUTHORS   Nauert,J.B., Klauck,T.K., Langeberg,L.K. and Scott,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1996) Biochemistry, OHSU, 3181 SW Sam Jackson
            Parkway, Portland, OR 97201, USA
COMMENT     On Jun 26, 1997 this sequence version replaced gi:1850362.
FEATURES             Location/Qualifiers
     source          1..6608
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             192..5537
                     /note="kinase scaffold protein"
                     /codon_start=1
                     /product="gravin"
                     /protein_id="AAC51366.1"
                     /db_xref="GI:2218077"
                     /translation="MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADP
                     AIAASDPATKLLQKNGQLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEV
                     IVTEVGQRDSEDVSERDSDKEMATKSAVVHDITDDGQEENRNIEQIPSSESNLEELTQ
                     PTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVKKDEGEGAAGAGDHQDPS
                     LGAGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAVEECKEEGEKQEKEP
                     SKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKV
                     DTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSE
                     EKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTEVEETAGSVPAEELVGMD
                     AEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERM
                     KVQGSPLKKLFTSTGLKKLSGKKQKGKRGGDEESGEHTQVPADSPDSQEEQKGESSA
                     SSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRR
                     PSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALI
                     CVGSSKKRARRRSSSDEEGGPKAMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGS
                     SSPEQAGSPTEGEGVSTWESFKRLVTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPG
                     KEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVER
                     EKMEAQQAQKGAEQPEQKAATEVSKELSESQVHMMAAAVADGTRAATIIEERSPSWIS
                     ASVTEPLEQVEAEAALLTEEVLEREVIAEEEPPTVTEPLPENREARGDTVVSEAELTP
                     EAVTAAETAGPLGSEEGTEASAAEETTEMVSAVSQLTDSPDTTEEATPVQEVEGGVPD
                     IEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDVLQPVQRAEAERPEEQAEASGLKKE
                     TDVVLKVDAQEAKTEPFTQGKVVGQTTPESFEKAPQVTESIESSELVTTCQAETLAGV
                     KSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIHEENEVASGTQ
                     SGGTEAEAVPAQKERPPAPSSFVFQEETKEQSKMEDTLEHTDKEVSVETVSILSKTEG
                     TQEADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDALE
                     LQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSKQLLQTVN
                     VPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLGETANI
                     LETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATTKKG
                     LSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEDLEPENGILELETKSSKLVQNI
                     IQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQETEKEGEEPQASAQDETPI
                     TSAKEESESTAVGQAHSDISKDMSEASEKTMTVEVEGSTVNDQQLEEVVLPSEEEGGG
                     AGTKSVPEDDGHALLAERIEKSLVEPKEDEKGDDVDDPENQNSALADTDASGGLTKES
                     PDTNGPKQKEKEDAQEVELQEGKVHSESDKAITPQAQEELQKQERESAKSELTES"
BASE COUNT     2091 a   1418 c   1827 g   1272 t
ORIGIN

  Query Match             34.8%;  Score 2144.4;  DB 9;  Length 6608;
  Best Local Similarity   64.3%;  Pred. No. 0;
  Matches 4215;  Conservative    0;  Mismatches 1796;  Indels  543;  Gaps   45;
```

```
Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
          |||  |   |  ||||| ||     |||||||||||| || ||  |||||||||||||
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
          ||||| ||||||||      | ||||||  |||||||    |   || ||  ||  |||
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
          || |||  | || || ||   ||  |          || |      | |  |||  |  | |||||
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
          ||||||||||||||| ||||||||||| ||||||||   |   ||||  ||||||||||
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
          ||||  ||  ||     |||| ||||    |||     |  ||
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy    269 -------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
                              | |||| |||||||| |||      ||| ||||||||| ||
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521
```

```
Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1461 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1496

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
```

```
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3833

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3834 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3893

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3894 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3953

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3954 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4013

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4014 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4073

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4074 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4133

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4134 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4193

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4194 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4253

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4254 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4313

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4314 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4373

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4374 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4433

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4434 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4493

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4494 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4553

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4554 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4613

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4614 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4673

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4674 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4733

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4734 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4784

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4785 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4844
```

```
Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4845 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4904

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4905 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4964

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4965 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5024

Qy   4721 GCAGCCCA---------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5025 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5084

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5085 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5144

Qy   4819 ------------------------------------------------------------ 4818
Db   5145 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5204

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5205 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5264

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5265 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5324

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5325 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5384

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5385 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5444

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5445 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5504

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5505 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5564

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5565 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5623

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5624 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5682

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5683 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5741

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5742 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5781

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5782 ------------------------------------------CATATAAATTCTGATTTA 5799

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5800 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5859

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5860 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5916

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5917 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5976

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5977 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6036

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6037 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6096

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6097 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6156

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6157 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6216

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6217 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6269

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6270 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6329

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6330 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6389

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6390 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6447

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6448 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6507

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6508 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6567

Qy   6134 CTGATTTAAATAAA 6147
Db   6568 CTGATTTAAATAAA 6581
```

RESULT 12
AF001504

```
LOCUS       AF001504            6297 bp    mRNA    linear   PRI 06-JUN-1997
DEFINITION  Homo sapiens myasthenia gravis autoantigen gravin mRNA, partial
            cds.
ACCESSION   AF001504
VERSION     AF001504.1  GI:2160783
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6297)
  AUTHORS   Bowditch,R.D. and Ginsberg,M.H.
  TITLE     Sequence of gravin cDNA isolated from a human umbilical vein
            endothelial cell library
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6297)
  AUTHORS   Bowditch,R.D. and Ginsberg,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1997) Biochemistry, University of Oklahoma Health
```

```
            Sciences Center, PO Box 26901, BMSB 853, Oklahoma City, OK 73190,
            USA
FEATURES             Location/Qualifiers
     source          1..6297
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="endothelial cell"
                     /tissue_type="umbilical vein"
     CDS             <1..5226
                     /note="kinase scaffold protein"
                     /codon_start=1
                     /product="myasthenia gravis autoantigen gravin"
                     /protein_id="AAB58938.1"
                     /db_xref="GI:2160784"
                     /translation="DPAIAASDPATKLLQKNGQLSTINGVAEQDELSLQEGDLNGQKG
                     ALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDSDKEMATKSAVVHDITDDGQEETPE
                     IIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVK
                     KDEGEGAAGAGDHQDPSLGAGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQ
                     AVEECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKE
                     DEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEK
                     VELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTE
                     VEETAGSVPAEELVEMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKP
                     PEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGDEESGEHTQV
                     PADSPDSQEEQKGESSASSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVT
                     PWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEPKP
                     EEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPKAMGGDHQKADEAGKDKETG
                     TDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRLVTPRKKSKSKLEEKSED
                     SIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPTGANEDD
                     SDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELSESQVHMMAAAVAD
                     GTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEEEPPTVTEPLPE
                     NREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSAVSQLTDSPD
                     TTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDVLQPVQRA
                     EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPESFEKAPQVTESI
                     ESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQK
                     DEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPAPSSFVFQEETKEQSKMEDTLEHT
                     DKEVSVETVSILSKTEGTQEADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVAL
                     KGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHET
                     AVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFT
                     LTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPD
                     CQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEEDLEP
                     ENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQE
                     TEKEGEEPQASAQDETPITSAKEESESTAVGQAHSDISKDMSEASEKTMTVEVEGSTV
                     NDQQLEEVVLPSEEEGGGAGTKSVPEDDGHALLAERIEKSLVEPKEDEKGDDVDDPEN
                     QNSALADTDASGGLTKESPDTNGPKQKEKEDAQEVELQEGKVHSESDKAITPQAQEEL
                     QKQERESAKSELTES"
BASE COUNT     2051 a   1310 c   1708 g   1228 t
ORIGIN

  Query Match             34.7%;  Score 2136.8;  DB 9;  Length 6297;
  Best Local Similarity   64.5%;  Pred. No. 0;
  Matches 4120;  Conservative    0;  Mismatches 1747;  Indels  525;  Gaps   41;

Qy     153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212
Db       7 GCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACC 66

Qy     213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGC-- 270
Db      67 ATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAG 126

Qy     271 ----------------------------------------AGGAGGAAGAAGTCGTTGAT 290
Db     127 AAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTC 186

Qy     291 GAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATG 350
Db     187 ACAGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATG 246

Qy     351 GCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAA 410
Db     247 GCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAA 306

Qy     411 ATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAG 470
Db     307 ATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAG 366

Qy     471 TCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACG 530
Db     367 TCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACT 426

Qy     531 GTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGAT 590
Db     427 GTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGAT 486

Qy     591 GAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAG 650
Db     487 GAAGGGGAGGGAGCAGCA---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGG 537

Qy     651 ACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAA 710
Db     538 GCT------GGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCC 591

Qy     711 GAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGAT 770
Db     592 GAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGC 651

Qy     771 CAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACC 830
Db     652 CAA---GCAGTGGAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGC 708

Qy     831 AAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAG 890
Db     709 AAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAA 768

Qy     891 TTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAG 950
Db     769 TTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAG 828

Qy     951 GATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAA 1010
Db     829 GATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAA 888

Qy    1011 GAAAAGGAAAAGACAGAGCCAGCCTCGGAGGA-------------------------G 1043
Db     889 GAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCA 948

Qy    1044 CAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAG 1103
Db     949 CAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAG 1008

Qy    1104 CTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCT 1163
Db    1009 CTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCG 1068

Qy    1164 TTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAG 1220
Db    1069 TTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAA 1128

Qy    1221 GTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAA 1280
Db    1129 GTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAGAAA------------------ 1170

Qy    1281 GGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCC 1340
Db    1171 ------ACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATG 1224

Qy    1341 CAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGT 1400
Db    1225 GATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGT 1284

Qy    1401 GTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCC 1460
Db    1285 GTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCC 1344

Qy    1461 AAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAG 1520
```

```
Db   1345 AAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAG 1404

Qy   1521 GTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGG 1580
Db   1405 GTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGA 1464

Qy   1581 AAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACAC 1640
Db   1465 AAGAAACAGAAAGGGAAAAGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAG 1521

Qy   1641 ATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCC 1700
Db   1522 GTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCC 1581

Qy   1701 CCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGAT 1760
Db   1582 CCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGAT 1641

Qy   1761 GGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCC 1820
Db   1642 GGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCC 1701

Qy   1821 TGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGT 1880
Db   1702 TGGGCTTCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGT 1761

Qy   1881 GACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACA 1940
Db   1762 GATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACA 1821

Qy   1941 GTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCA 2000
Db   1822 GCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCA 1881

Qy   2001 AAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAG 2060
Db   1882 AAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAG 1941

Qy   2061 AAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGG 2120
Db   1942 AAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGA 2001

Qy   2121 GACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCT 2180
Db   2002 GACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTT 2061

Qy   2181 GCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGC 2240
Db   2062 GCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGC 2121

Qy   2241 CCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGA 2300
Db   2122 CCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGA 2181

Qy   2301 AAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTA 2348
Db   2182 AAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTA 2241

Qy   2349 GAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAA 2408
Db   2242 GAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAG 2301

Qy   2409 TTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAA 2468
Db   2302 TTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAA 2361

Qy   2469 GACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTG 2528
Db   2362 GACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTG 2421

Qy   2529 TCTGAGTATAATGCAGTGGAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACG 2579
Db   2422 TCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCA 2481

Qy   2580 GAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTC 2639
Db   2482 GAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTT 2541

Qy   2640 CACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGG 2699
Db   2542 CATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGG 2601

Qy   2700 TCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCC 2759
Db   2602 TCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCC 2661

Qy   2760 ATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTG 2816
Db   2662 GCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACG 2721

Qy   2817 CTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTG 2876
Db   2722 GTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCG 2781

Qy   2877 GATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAA 2936
Db   2782 GAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAA 2841

Qy   2937 GTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTG 2996
Db   2842 GGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTA 2901

Qy   2997 ACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTG 3056
Db   2902 ACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTA 2961

Qy   3057 CTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAG 3116
Db   2962 CCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAA 3021

Qy   3117 GTGAAAGAGGAGTCCCAGGTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAA 3173
Db   3022 GTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTG 3081

Qy   3174 GCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAG 3233
Db   3082 CAGAGAGCAGAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAG 3141

Qy   3234 AAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGC 3293
Db   3142 ACGGATGTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGG 3201

Qy   3294 TCTGAGACTGGACAGGCTACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGAT 3350
Db   3202 AAGGTGGTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGC 3261

Qy   3351 GTAGAC---------CATGTCGCCACGTGCCA-----------------GGTTATCAAG 3383
Db   3262 ATAGAGTCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCA 3321

Qy   3384 CTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGT 3443
Db   3322 CAGGAGATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGT 3381

Qy   3444 GAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGAT 3503
Db   3382 GAGACTGATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGAC 3441

Qy   3504 GAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACA 3563
Db   3442 GAGATTGTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACA 3501

Qy   3564 GAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCA 3623
Db   3502 GAAGCAGAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTC 3561
```

```
Qy   3624 GCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAG 3683
Db   3562 CAGGAAGAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAG 3621
Qy   3684 CTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTT 3743
Db   3622 GTGTCAGTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAG 3681
Qy   3744 GACTGGTTGGATGGAGAAAAAGTCAAA---------GAAGAACAGGAGGTGTTTGTACAC 3794
Db   3682 TATGCTGATGAGAAAACCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGAC 3741
Qy   3795 TCTGG------ACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATG 3848
Db   3742 ACAGGCATAACAGTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACA 3801
Qy   3849 GGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAG---------- 3898
Db   3802 GAAGAAGCTGAATGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCT 3861
Qy   3899 -----CCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAG 3953
Db   3862 CCATCCCCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAG 3921
Qy   3954 CCAGAGCAAGTGAGTGAAG---AAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGA 4010
Db   3922 CCAACCCATGTGAATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAG 3981
Qy   4011 ACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTG 4070
Db   3982 GTCAGTAAGCAGCTCCTCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTC 4041
Qy   4071 GGAAGCCTTGGAGGAAGCCCTTCTCT---CCCAGACCAAGACAAAGCAGGTTGCATAGAG 4127
Db   4042 AGCAGTTTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAA 4101
Qy   4128 GTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA---- 4183
Db   4102 ATTCAAGTTCAGAGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAG 4161
Qy   4184 --GGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCA 4241
Db   4162 GTCTTAGGAGAAACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCA 4221
Qy   4242 CACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCA 4301
Db   4222 CATTTAGTTCTGGAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGG 4281
Qy   4302 GAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACT 4361
Db   4282 GAAGATGCTGTGCCCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCT 4341
Qy   4362 CCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGA 4421
Db   4342 GCTACTACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAG 4401
Qy   4422 GAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGT 4481
Db   4402 TGGAAGTCAGATGAAGTCGATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGT 4455
Qy   4482 ACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGC 4541
Db   4456 GTAGCAATTGAGGAGGATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGC 4515
Qy   4542 AACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAA 4598
Db   4516 AGTAAACTTGTCCAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAA 4575
Qy   4599 ACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTC----------------C 4642
Db   4576 ACAGCCACCGAAATGTTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGAC 4635
Qy   4643 TGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAA 4702
Db   4636 AGCCAGGACGCTGGACAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAG 4695
Qy   4703 GATGAAACACCCAGTGCCGCAGCCCA--------------------GAGAGGACTTGCA 4741
Db   4696 GATGAAACACCAATTACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCA 4755
Qy   4742 AGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTT 4801
Db   4756 CATTCTGATATTTCCAAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTA 4815
Qy   4802 GAAAGCGCCGGTGTCAA------------------------------------------- 4818
Db   4816 GAAGGTTCCACTGTAAATGATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAG 4875
Qy   4819 --------------------------------------------------------AGTA 4822
Db   4876 GGAGGTGGAGCTGGAACAAAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAA 4935
Qy   4823 AGCATTGAGAAGCTGCCTCCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGC 4880
Db   4936 AGAATAGAGAAGTCACTAGTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGAC 4995
Qy   4881 CCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAA 4940
Db   4996 CCTGAAAACCAGAACTCAGCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAG 5055
Qy   4941 TCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTC 5000
Db   5056 TCCCCAGATACAAATGGACCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTG 5115
Qy   5001 CAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGAC 5060
Db   5116 CAGGAAGGAAAAGTGCACAGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAG 5175
Qy   5061 CTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTG------------- 5107
Db   5176 TTACAGAAACAAGAGAGAGAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATG 5235
Qy   5108 --------CTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAG-- 5157
Db   5236 CAGTTAAACTCATTGTCTGTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAAT 5295
Qy   5158 --ATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAG 5215
Db   5296 GAATGCTGCTGCTGAGA-CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAG 5353
Qy   5216 GGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAG 5275
Db   5354 GCACATCAACTGATCTCATTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAG 5412
Qy   5276 CTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGT 5335
Db   5413 CTAGAGCCATTTAACATTTCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-- 5470
Qy   5336 CCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGT 5395
Db   5471 ------------------------------------------------------------ 5470
Qy   5396 CCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCT 5454
Db   5471 CATATAAATCTGATTTAAGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCT 5530
Qy   5455 AGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAG 5514
Db   5531 AGTTCTGCTTCTGAAACTGGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG 5590
Qy   5515 TCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTT----------------AAG 5557
Db   5591 ---TCCTCCTGTATCTATTGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATAC 5647
Qy   5558 CACATGCTTTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAA 5617
```

```
Db   5648 TACATGCTTTTTGTATCACACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGG 5707

Qy   5618 AGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAGATAACATTC---------CTGGCAA 5668
Db   5708 AGCTTTAAGCCTCAGTTATATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCT 5767

Qy   5669 GAAGAGACAAGTCTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCT 5728
Db   5768 GATCAAGGTACAATTCTTTAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCT 5827

Qy   5729 GAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACAT 5788
Db   5828 GAAATTGGTCACTTTCCTATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACAT 5887

Qy   5789 CCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTT 5848
Db   5888 ACAGAATGTTCTATTGTCATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTT 5940

Qy   5849 GGAAAGAAGTCTCCTCTGTGTCAGA-----------------TTAAGCCCTGTCTCTTAA 5891
Db   5941 AGAAAGAAGTTATATTCTGGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGG 6000

Qy   5892 TGATATGGACAAATGAGTGTGCCTAAGGCCATGAGA----------------TGTTTCCT 5935
Db   6001 TTGTTTGGACCGATAAGTGTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGT 6060

Qy   5936 AATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATT 5995
Db   6061 TATGAAGAAAAGAATTGTTGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATT 6118

Qy   5996 CATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGT 6055
Db   6119 CACACATGGCATGAAATAAGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATT 6178

Qy   6056 TTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGA 6115
Db   6179 GTGTTTGTGCCATATTTGTGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGA 6238

Qy   6116 TAAATTGTGATTTGACAACTGATTTAAATAAA 6147
Db   6239 TAAGTTGTGATTTGACGACTGATTTAAATAAA 6270


RESULT 13
AR002603
LOCUS       AR002603                6605 bp    DNA     linear   PAT 04-DEC-1998
DEFINITION  Sequence 4 from patent US 5741890.
ACCESSION   AR002603
VERSION     AR002603.1  GI:3964157
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6605)
  AUTHORS   Scott,J.D., Nauert,J.Brian. and Klauck,T.M.
  TITLE     Protein binding fragments of gravin
  JOURNAL   Patent: US 5741890-A 4 21-APR-1998;
FEATURES             Location/Qualifiers
     source          1..6605
                     /organism="unknown"
BASE COUNT     2091 a   1416 c   1826 g   1272 t
ORIGIN

  Query Match             34.4%;  Score 2118.8;  DB 6;  Length 6605;
  Best Local Similarity   64.2%;  Pred. No. 0;
  Matches 4206;  Conservative    0;  Mismatches 1802;  Indels  546;  Gaps   46;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy    269 -------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy   1029 CCAGCCTCGGAGGA-------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340
```

```
Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1461 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1496

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
```

```
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGA---GGTGCATCTGGTACCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCT 3830

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3831 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTGTGTTCCAGGAAGAAACTAAAGAA 3890

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3891 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3950

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3951 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4010

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4011 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4070

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4071 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4130

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4131 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4190

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4191 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4250

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4251 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4310

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4311 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4370

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4371 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4430

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4431 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4490

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4491 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4550

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4551 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4610

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4611 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4670

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4671 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4730

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4731 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4781

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4782 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4841

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4842 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4901

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4902 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4961

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4962 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5021

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5022 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5081

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5082 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5141

Qy   4819 ------------------------------------------------------------ 4818
Db   5142 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5201

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5202 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5261

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5262 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5321

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5322 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5381

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5382 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5441

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5442 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5501

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5502 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5561

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5562 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5620

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5621 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5679
```

```
Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5680 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5738

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5739 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5778

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5779 -----------------------------------------CATATAAATTCTGATTTA 5796

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5797 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5856

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5857 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5913

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5914 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5973

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5974 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6033

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6034 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6093

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6094 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6153

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6154 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6213

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6214 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6266

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6267 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6326

Qy   5910 GTGCCTAAGGCCATGAGA---------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6327 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6386

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6387 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6444

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6445 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6504

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6505 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6564

Qy   6134 CTGATTTAAATAAA 6147
Db   6565 CTGATTTAAATAAA 6578

RESULT 14
AB003476
LOCUS       AB003476                6287 bp    mRNA    linear   PRI 05-FEB-1999
DEFINITION  Homo sapiens mRNA for gravin, complete cds.
ACCESSION   AB003476
VERSION     AB003476.1  GI:2081606
KEYWORDS    gravin.
SOURCE      Homo sapiens Umbilical Vein Endothelial Cell cDNA to mRNA.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (sites)
  AUTHORS   Sato,N., Kokame,K., Shimokado,K., Kato,H. and Miyata,T.
  TITLE     Changes of cell growth-related and other gene expression by
            atherogenic lipid, lysophosphatidylcholine, in cultured human
            umbilical vein endothelial ce lls
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6287)
  AUTHORS   Kokame,K.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-APR-1997) Koichi Kokame, National Cardiovascular
            Center Research Institute, Department of Etiology and Pathogenesis;
            Fujishirodai 5-7-1, Suita, Osaka 565, Japan
            (E-mail:kame@ri.ncvc.go.jp, Tel:+81-6-833-5012, Fax:+81-6-872-8091)
FEATURES             Location/Qualifiers
     source          1..6287
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="Endothelial Cell"
                     /tissue_type="Umbilical Vein"
     CDS             174..5228
                     /codon_start=1
                     /product="gravin"
                     /protein_id="BAA19927.1"
                     /db_xref="GI:2081607"
                     /translation="MLGTITITVGQRDSEDVSKRDSDKEMATKSAVVHDITDDGQEET
                     PEIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLT
                     VKKDEGEGAAGAGDHKDPSLGAGEAASKESEPKQSTEKPEETLKREQSHAEISPPAES
                     GQAVEECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKP
                     KEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEY
                     EKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQK
                     TEVEETAGSVPAEELVEMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLS
                     KPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGGDEESGEHT
                     QVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREG
                     VTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEP
                     KPEEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPKAMGGDHQKADEAGKDKE
                     TGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRLVTPRKKSKSKLEEKS
                     EDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPTGANE
                     DDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELSESQVHMMAAAV
                     ADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEEEPPTVTEPL
                     PENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSAVSQLTDS
                     PDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDVLQPVQ
                     RAEAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPESFEKAPQVTE
                     SIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTT
                     QKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPAPSSFVFQEETKEQSKMEDTLE
                     HTDKEVSVETVSILSKTEGTQEADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEV
                     ALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEH
                     ETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEAS
                     FTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG
                     PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEDLE
                     PENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQ
                     ETEKEGEEPLASAQDETPITSAKEESESTAVGQAHSDISKDMSEASEKTMTVEVEGST
                     VNDQQLEEVVLPSEEEGGGAGTKSVPEDDGHALLAERIEKSLVEPKEDEKGDDVDDPE
                     NQNSALADTDASGGLTKESPDTNGPKQKEKEDAQEVELQEGKVHSESDKAITPQAQEE
                     LQKQERESAKSELTES"
BASE COUNT     2021 a   1313 c   1710 g   1243 t
ORIGIN

  Query Match             33.9%;  Score 2089.6;  DB 9;  Length 6287;
  Best Local Similarity   64.6%;  Pred. No. 0;
  Matches 4011;  Conservative    0;  Mismatches 1709;  Indels  486;  Gaps   41;

Qy    297 GTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAGACCGAGTTGAAGAAATGGCGGCC 356
Db    198 GTTGGACAGAGAGACTCTGAAGATGTGAGCAAAAGAGACTCCGATAAAGAGATGGCTACT 257
```

```
Qy    357 AACTCCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATT 416
Db    258 AAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATC 317

Qy    417 GAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAG 476
Db    318 GAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAG 377

Qy    477 GCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAG 536
Db    378 GCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAA 437

Qy    537 AAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGC 596
Db    438 AAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGG 497

Qy    597 GAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCC 656
Db    498 GAGGGAGCAGCA---------GGGGCTGGCGACCACAAGGACCCCAGCCTTGGGGCT--- 545

Qy    657 GTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGC 716
Db    546 ---GGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAG 602

Qy    717 ACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCG 776
Db    603 ACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAA--- 659

Qy    777 GCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCC 836
Db    660 GCAGTGGAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCT 719

Qy    837 CCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTC 896
Db    720 GCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTC 779

Qy    897 ACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGAT 956
Db    780 ACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAA 839

Qy    957 CTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAG 1016
Db    840 GTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGAC 899

Qy   1017 GAAAAGACAGAGCCAGCCTCGGAGGA--------------------------GCAGGAG 1049
Db    900 GGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAG 959

Qy   1050 CCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCT 1109
Db    960 CCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCC 1019

Qy   1110 TTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCA 1169
Db   1020 TCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCG 1079

Qy   1170 ACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCAC 1226
Db   1080 ACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCAC 1139

Qy   1227 GTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGC 1286
Db   1140 GTCAGCACCGTGGAGGAGAGAACCGAAGAGCAGAAA----------------------- 1175

Qy   1287 GTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAG 1346
Db   1176 ACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCA 1235

Qy   1347 GTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCT 1406
Db   1236 GAACCTCAGGAAGCTGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCC 1295

Qy   1407 GGAGGAGACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACAC 1466
Db   1296 GGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCC 1355

Qy   1467 CCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAG 1526
Db   1356 CCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAG 1415

Qy   1527 GGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAG 1586
Db   1416 GGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAA 1475

Qy   1587 CAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCAC 1646
Db   1476 CAGAAAGGGAAAAGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCA 1532

Qy   1647 ACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAG 1706
Db   1533 GCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAG 1592

Qy   1707 GAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAA 1766
Db   1593 GAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAA 1652

Qy   1767 GCTGAGGAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCA 1826
Db   1653 GCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCA 1712

Qy   1827 TCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAG 1886
Db   1713 TCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAA 1772

Qy   1887 GAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCA 1946
Db   1773 GAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCT 1832

Qy   1947 GAAATGCAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGT 2006
Db   1833 GAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGC 1892

Qy   2007 AGGGTGGATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGA 2066
Db   1893 AAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGA 1952

Qy   2067 GCAAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGT 2126
Db   1953 GCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCAC 2012

Qy   2127 CACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGC 2186
Db   2013 CAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGATCCTTGCTGGT 2072

Qy   2187 ACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCC 2246
Db   2073 TCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACC 2132

Qy   2247 GAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAA 2306
Db   2133 GAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAA 2192

Qy   2307 TCCAAGTCAAAACTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAG 2354
Db   2193 TCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACAT 2252

Qy   2355 TTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATC 2414
Db   2253 TCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATT 2312

Qy   2415 CCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCA 2474
Db   2313 CCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCA 2372

Qy   2475 GGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAG 2534
```

```
Db   2373 GGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAG 2432

Qy   2535 TATAATGCAGTGGAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTG 2585
Db   2433 TATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAG 2492

Qy   2586 CCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACT 2645
Db   2493 CCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATG 2552

Qy   2646 GTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCT 2705
Db   2553 ATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCT 2612

Qy   2706 TCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCA 2765
Db   2613 TCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTG 2672

Qy   2766 CCTGTTGAAGAGGTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACC 2822
Db   2673 TTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACT 2732

Qy   2823 CAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTC 2882
Db   2733 GAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTG 2792

Qy   2883 ACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACC 2942
Db   2793 ACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACC 2852

Qy   2943 GAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGAC 3002
Db   2853 GAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGAC 2912

Qy   3003 TCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGAT 3062
Db   2913 TCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGAC 2972

Qy   3063 ACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAA 3122
Db   2973 ATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAA 3032

Qy   3123 GAGGAGTCCCAGGTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTG 3179
Db   3033 GAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGA 3092

Qy   3180 GAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGAC 3239
Db   3093 GCAGAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGAT 3152

Qy   3240 GTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAG 3299
Db   3153 GTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTG 3212

Qy   3300 ACTGGACAGGCTACTCCAGAGAGCCTTGAA---GTTCCTGAAGTCACGGCAGATGTAGAC 3356
Db   3213 GTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAG 3272

Qy   3357 ---------CATGTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAG 3389
Db   3273 TCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAG 3332

Qy   3390 CAGCTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACA 3449
Db   3333 ATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACT 3392

Qy   3450 AATGGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACC 3509
Db   3393 GATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATT 3452

Qy   3510 ATTGACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAA 3569
Db   3453 GTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCA 3512

Qy   3570 GAGGCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAG 3629
Db   3513 GAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAA 3572

Qy   3630 GAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACT 3689
Db   3573 GAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCA 3632

Qy   3690 GCTGCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGG 3749
Db   3633 GTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCT 3692

Qy   3750 TTGGATGGAGAAAAAGTCAAA---------GAAGAACAGGAGGTGTTTGTACACTCTGG- 3799
Db   3693 GATGAGAAAACCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGC 3752

Qy   3800 -----ACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTG 3854
Db   3753 ATAACAGTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTAAAGGTGAAGGGACAGAAGAA 3812

Qy   3855 GCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAG--------------C 3899
Db   3813 GCTGAATGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCC 3872

Qy   3900 CTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAG 3959
Db   3873 CCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACC 3932

Qy   3960 CAAGTGAGTGAAG---AAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTAC 4016
Db   3933 CATGTGAATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGT 3992

Qy   4017 GGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGC 4076
Db   3993 AAGCAGCTCCTCCAGACAGTGAATGTGCCCATCATAGATGGAGCAAAGGAAGTCAGCAGT 4052

Qy   4077 CTTGGAGGAAGCCCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAA 4133
Db   4053 TTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAA 4112

Qy   4134 GTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTC 4187
Db   4113 GTTCAGAGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTA 4172

Qy   4188 ATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTA 4247
Db   4173 GGAGAAACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTA 4232

Qy   4248 TTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGAC 4307
Db   4233 GTTCTGGAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGAT 4292

Qy   4308 ACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCT 4367
Db   4293 GCTGTGCCCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACT 4352

Qy   4368 CCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGA 4427
Db   4353 ACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAG 4412

Qy   4428 TCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCA 4487
Db   4413 TCAGATGAAGTCGATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCA 4466

Qy   4488 ATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAG 4547
Db   4467 ATTG---AGGATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAA 4523

Qy   4548 ATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCC 4604
Db   4524 CTTGTCCAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCC 4583
```

```
Qy   4605 CCCGAAACTCATGCTTATGATTCACAGACCCAGGTTC----------------CTGCATG 4648
Db   4584 ACCGAAATGTTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAG 4643

Qy   4649 CAGGCTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAA 4708
Db   4644 GACGCTGGACAGGAAACGGAGAAAGAAGGAGAGGAACCTCTGGCCTCTGCACAGGATGAA 4703

Qy   4709 ACACCCAGTGCCGCAGCCCA--------------------GAGAGGACTTGCAAGTCCT 4747
Db   4704 ACACCAATTACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCT 4763

Qy   4748 GACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGC 4807
Db   4764 GATATTTCCAAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGT 4823

Qy   4808 GCCGGTGTCAA------------------------------------------------- 4818
Db   4824 TCCACTGTAAATGATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGT 4883

Qy   4819 --------------------------------------------------AGTAAGCATT 4828
Db   4884 GGAGCTGGAACAAAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATA 4943

Qy   4829 GAGAAGCTGCCTCCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAG 4886
Db   4944 GAGAAGTCACTAGTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAA 5003

Qy   4887 CTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCA 4946
Db   5004 AACCAGAACTCAGCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCA 5063

Qy   4947 GACACCAACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAA 5006
Db   5064 GATACAAATGGACCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAA 5123

Qy   5007 GAAGAAATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAG 5066
Db   5124 GGAAAAGTGCACAGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAG 5183

Qy   5067 GAGCCAAAGGGAGACCTGGCAGAATCCTAAGATGTTAGTT-------------------- 5106
Db   5184 AAACAAGAGAGAGAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTA 5243

Qy   5107 -GCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGC 5161
Db   5244 AACTCATTGTCTGTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGC 5303

Qy   5162 TGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGT 5221
Db   5304 TGCTGCTGAGA-CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACAT 5361

Qy   5222 CCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAG 5281
Db   5362 CAACTGATCTCATTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAG 5420

Qy   5282 CCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGT 5341
Db   5421 CCATTTAACATTTCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------- 5472

Qy   5342 ATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGT 5401
Db   5473 ------------------------------------------------------CATATA 5478

Qy   5402 ATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCA-ATTCTGCCATACCTAGTTCC 5460
Db   5479 AATTCTGATTTAAGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCT 5538

Qy   5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520
Db   5539 GCTTCTGAAACTGGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCC 5595

Qy   5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTT-----------------AAGCACATG 5563
Db   5596 TCCTGTATCTATTGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATG 5655

Qy   5564 CTTTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCC 5623
Db   5656 CTTTTTGTATCACACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTT 5715

Qy   5624 AAGCCTCAACTGTAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAG 5674
Db   5716 AAGCCTCAGTTATATAACCCACGAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAA 5775

Qy   5675 ACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGT 5734
Db   5776 GGTACAATTCTTTAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATT 5835

Qy   5735 GGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGA 5794
Db   5836 GGTCACTTTCCTATTACACGGAGTGTGCTAAAACTAAAAAGCATTTTGAAACATACAGAA 5895

Qy   5795 TGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAG 5854
Db   5896 TGTTCTATTGTCATTGGGAAATTT-------TTCTTTCTAACCCAGTGGAGGTTAGAAAG 5948

Qy   5855 AAGTCTCCTCTGTGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATAT 5897
Db   5949 AAGTTATATTCTGGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTT 6008

Qy   5898 GGACAAATGAGTGTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCA 5941
Db   6009 GGACCGATAAGTGTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAA 6068

Qy   5942 GAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATG 6001
Db   6069 GAAAAGAATTGTTGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACA 6126

Qy   6002 CAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCT 6061
Db   6127 TGGCATGAAATAAGTCAGGTTCTTTACAAATGGTATTTGATAGATACTGGATTGTGTTT 6186

Qy   6062 GTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATT 6121
Db   6187 GTGCCATATTTGTGCCATTCTTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTT 6246

Qy   6122 GTGATTTGACAACTGATTTAAATAAA 6147
Db   6247 GTGATTTGACGACTGATTTAAATAAA 6272
```

```
RESULT 15
AL590413
LOCUS       AL590413              104939 bp    DNA     linear   PRI 06-DEC-2001
DEFINITION  Human DNA sequence from clone RP11-351K16 on chromosome 6, complete
            sequence.
ACCESSION   AL590413
VERSION     AL590413.18  GI:17426497
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104939)
  AUTHORS   Martin,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 8, 2001 this sequence version replaced gi:17384482.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
```

```
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-351K16 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-351K16 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-351K16 is at 104939 in this
            sequence. The true left end of clone RP11-108N8 is at 59858 in this
            sequence. The true right end of clone RP3-403M6 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1. .104939
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-351K16"
                     /clone_lib="RPCI-11.2"
     misc_feature    51026. .51045
                     /note="Sequence from AC023201 sequenced by WIBR."
     misc_feature    51799. .51952
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    52345. .52447
                     /note="Sequence from reads from a short insert library
                     derived from a single pUC clone.  Restriction digest data
                     confirm the assembly."
     misc_feature    86009. .86146
                     /note="Sequence from overlapping clone RP11-108N8
                     (AL590543).  Assembly confirmed by restriction digest."
     misc_feature    87770. .87807
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    30065 a  22273 c  22882 g  29719 t
ORIGIN

  Query Match            30.9%;  Score 1906.4;  DB 9;  Length 104939;
  Best Local Similarity  68.2%;  Pred. No. 0;
  Matches 3070;  Conservative    0;  Mismatches 1246;  Indels  186;  Gaps   23;

Qy      297 GTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAGACCGAGTTGAAGAAATGGCGGCC 356
Db     9415 GTTGGACAGAGAGACTCTGAAGATGTGAGCAAAAGAGACTCCGATAAAGAGATGGCTACT 9474

Qy      357 AACTCCACAGCTGTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATT 416
Db     9475 AAGTCAGCGGTTGTTCACGACATCACAGATGATGGCAGGAGGAGACACCCGAAATAATC 9534

Qy      417 GAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAG 476
Db     9535 GAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAG 9594

Qy      477 GCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAG 536
Db     9595 GCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAA 9654

Qy      537 AAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGC 596
Db     9655 AAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGG 9714

Qy      597 GAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCC 656
Db     9715 GAGGGAGCAGCA---------GGGGCTGGCGACCACAAGGACCCCAGCCTTGGGGCT--- 9762

Qy      657 GTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGC 716
Db     9763 ---GGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAG 9819

Qy      717 ACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCG 776
Db     9820 ACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAA--- 9876

Qy      777 GCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCC 836
Db     9877 GCAGTGGAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCT 9936

Qy      837 CCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTC 896
Db     9937 GCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTC 9996

Qy      897 ACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGAT 956
Db     9997 ACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAA 10056

Qy      957 CTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAG 1016
Db    10057 GTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGAC 10116

Qy     1017 GAAAAGACAGAGCCAGCCTCGGAGGA--------------------------GCAGGAG 1049
Db    10117 GGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAG 10176

Qy     1050 CCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCT 1109
Db    10177 CCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCC 10236

Qy     1110 TTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCA 1169
Db    10237 TCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCG 10296

Qy     1170 ACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCAC 1226
Db    10297 ACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCAC 10356

Qy     1227 GTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGC 1286
Db    10357 GTCAGCACCGTGGAGGAGAGAACCGAAGAGCAGAAA------------------------ 10392

Qy     1287 GTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAG 1346
Db    10393 ACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCA 10452

Qy     1347 GTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCT 1406
Db    10453 GAACCTCAGGAAGCTGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCC 10512

Qy     1407 GGAGGAGACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACAC 1466
Db    10513 GGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCC 10572

Qy     1467 CCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAG 1526
Db    10573 CCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAG 10632

Qy     1527 GGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAG 1586
Db    10633 GGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAA 10692

Qy     1587 CAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCAC 1646
```

```
Db  10693 CAGAAAGGGAAAAGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCA 10749

Qy   1647 ACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAG 1706
Db  10750 GCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAG 10809

Qy   1707 GAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAA 1766
Db  10810 GAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAA 10869

Qy   1767 GCTGAGGAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCA 1826
Db  10870 GCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCA 10929

Qy   1827 TCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAG 1886
Db  10930 TCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAA 10989

Qy   1887 GAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCA 1946
Db  10990 GAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCT 11049

Qy   1947 GAAATGCAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGT 2006
Db  11050 GAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGC 11109

Qy   2007 AGGGTGGATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGA 2066
Db  11110 AAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGA 11169

Qy   2067 GCAAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGT 2126
Db  11170 GCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCAC 11229

Qy   2127 CACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGC 2186
Db  11230 CAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGT 11289

Qy   2187 ACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCC 2246
Db  11290 TCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACC 11349

Qy   2247 GAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAA 2306
Db  11350 GAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAA 11409

Qy   2307 TCCAAGTCAAAACTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAG 2354
Db  11410 TCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACAT 11469

Qy   2355 TTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATC 2414
Db  11470 TCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATT 11529

Qy   2415 CCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCA 2474
Db  11530 CCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCA 11589

Qy   2475 GGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAG 2534
Db  11590 GGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAG 11649

Qy   2535 TATAATGCAGTGGAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTG 2585
Db  11650 TATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAG 11709

Qy   2586 CCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACT 2645
Db  11710 CCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATG 11769

Qy   2646 GTGAGTGTCGCAGTCATTGATGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCT 2705
Db  11770 ATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCT 11829

Qy   2706 TCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCA 2765
Db  11830 TCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTG 11889

Qy   2766 CCTGTTGAAGAGGTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACC 2822
Db  11890 TTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACT 11949

Qy   2823 CAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTC 2882
Db  11950 GAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTG 12009

Qy   2883 ACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACC 2942
Db  12010 ACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACC 12069

Qy   2943 GAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGAC 3002
Db  12070 GAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGAC 12129

Qy   3003 TCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGAT 3062
Db  12130 TCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGAC 12189

Qy   3063 ACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAA 3122
Db  12190 ATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAA 12249

Qy   3123 GAGGAGTCCCAGGTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTG 3179
Db  12250 GAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGA 12309

Qy   3180 GAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGAC 3239
Db  12310 GCAGAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGAT 12369

Qy   3240 GTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAG 3299
Db  12370 GTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTG 12429

Qy   3300 ACTGGACAGGCTACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTA--- 3353
Db  12430 GTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAG 12489

Qy   3354 ------GACCATGTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAG 3389
Db  12490 TCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAG 12549

Qy   3390 CAGCTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACA 3449
Db  12550 ATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACT 12609

Qy   3450 AATGGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACC 3509
Db  12610 GATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATT 12669

Qy   3510 ATTGACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAA 3569
Db  12670 GTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCA 12729

Qy   3570 GAGGCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAG 3629
Db  12730 GAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAA 12789

Qy   3630 GAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACT 3689
Db  12790 GAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCA 12849

Qy   3690 GCTGCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAG--------- 3740
Db  12850 GTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCT 12909
```

```
Qy     3741 GTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG- 3799

Db    12910 GATGAGAAAACCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGC 12969

Qy     3800 -----ACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTG 3854

Db    12970 ATAACAGTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAA 13029

Qy     3855 GCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGC 3899

Db    13030 GCTGAATGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCC 13089

Qy     3900 CTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAG 3959

Db    13090 CCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACC 13149

Qy     3960 CAAGTGAGTGAAG---AAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTAC 4016

Db    13150 CATGTGAATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGT 13209

Qy     4017 GGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGAAGGCACTGGGAAGC 4076

Db    13210 AAGCAGCTCCTCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGT 13269

Qy     4077 CTTGGAGGAAGCCCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAA 4133

Db    13270 TTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAA 13329

Qy     4134 GTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCAT---- 4189

Db    13330 GTTCAGAGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTA 13389

Qy     4190 --AGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTA 4247

Db    13390 GGAGAAACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTA 13449

Qy     4248 TTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGAC 4307

Db    13450 GTTCTGGAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGAT 13509

Qy     4308 ACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCT 4367

Db    13510 GCTGTGCCCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACT 13569

Qy     4368 CCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGA 4427

Db    13570 ACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAG 13629

Qy     4428 TCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCA 4487

Db    13630 TCAGATGAAGTCGATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCA 13683

Qy     4488 ATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAG 4547

Db    13684 ATTG---AGGATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAA 13740

Qy     4548 ATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCC 4604

Db    13741 CTTGTCCAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCC 13800

Qy     4605 CCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGG 4664

Db    13801 ACCGAAATGTTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAG 13860

Qy     4665 GA 4666
            ||
Db    13861 GA 13862
```

Search completed: December 12, 2002, 21:09:01
Job time : 16311 secs

```
OM nucleic - nucleic search, using sw model

Run on:         December 12, 2002, 14:13:30 ; Search time 1074 Seconds
                                              (without alignments)
                                              12916.489 Million cell updates/sec

Title:          US-09-902-432-3
Perfect score:  6160
Sequence:       1 cgggccggcgggagtagaag..........aaataaacatatttgactac 6160

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters:      4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      N_Geneseq_101002:*
                1:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                2:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                3:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                4:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                5:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                6:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                7:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                8:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                9:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                10:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                11:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                12:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                13:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                14:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                15:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                16:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                17:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                18:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                19:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                20:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                21:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                22:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                23:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                24:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	5182.4	84.1	5200	18	AAV02302	Rat tumour suppres
	2	4696.6	76.2	5074	18	AAV02301	Rat tumour suppres
	3	2179.6	35.4	6614	24	ABN59861	Novel human coding
	4	2144.4	34.8	6608	24	ABN96942	Gene #3440 used to
	5	2144.4	34.8	6608	24	ABL61938	Colon adenocarcino
	6	2118.8	34.4	6605	19	AAV23545	Human gravin polyp
	7	2118.8	34.4	6605	21	AAA74903	Human gravin codin
	8	1920.6	31.2	6886	23	AAS85205	DNA encoding novel
c	9	550	8.9	563	24	ABK63316	Rat sequence diffe
	10	172.2	2.8	377	20	AAX40328	Human secreted pro
	11	162.2	2.6	178	24	ABK63568	Rat sequence diffe
	12	105.2	1.7	483	20	AAV86620	EST clone BF146.
c	13	101.8	1.7	459	23	ABV56830	Human prostate exp
	14	95.4	1.5	400	22	AAI86887	Human polynucleoti
	15	92	1.5	339	20	AAV86330	EST clone AJ169.
	16	81.8	1.3	10732	21	AAA10594	Gene encoding a su
	17	79.4	1.3	385	22	AAI83958	Human polynucleoti
	18	78.6	1.3	3399	17	AAT05868	Chicken leucocytoz
	19	76	1.2	49999	20	AAZ23891	Murine LOBO genomi
	20	76	1.2	49999	20	AAZ23896	Murine LOBO homolo
	21	72.8	1.2	2800	22	AAD07131	Canine retinitis p
	22	72.4	1.2	1686	16	AAQ87587	DNA encoding Leuco
	23	72.2	1.2	6741	21	AAA10595	Gene encoding a su
	24	71	1.2	575	22	ABA50472	Human breast cell
	25	71	1.2	575	22	ABA68422	Human foetal liver
	26	71	1.2	575	22	ABA35413	Probe #13879 for g
	27	71	1.2	575	22	AAK16793	Human brain expres
	28	71	1.2	575	22	AAK42567	Human bone marrow
	29	71	1.2	575	22	AAI23315	Probe #13248 for g
	30	71	1.2	575	22	AAI48636	Probe #17322 used
	31	71	1.2	575	22	AAI08956	Probe #8947 used t
	32	71	1.2	575	24	ABS16615	Human genome-deriv
	33	71	1.2	1969	22	ABA45341	Human breast cell
	34	71	1.2	1969	22	ABA55830	Human foetal liver
	35	71	1.2	1969	22	ABA25506	Probe #3972 for ge
	36	71	1.2	1969	22	AAK04048	Human brain expres
	37	71	1.2	1969	22	AAK29533	Human bone marrow
	38	71	1.2	1969	22	AAI14105	Probe #4038 for ge
	39	71	1.2	1969	22	AAI35486	Probe #4172 used t
	40	71	1.2	1969	22	AAI03958	Probe #3949 used t
	41	71	1.2	1969	24	ABS04084	Human genome-deriv
	42	70.8	1.1	799	19	AAV55831	Nucleotide sequenc
	43	70.8	1.1	1926	21	AAA50254	Epstein Barr virus
	44	70.8	1.1	1926	22	AAF82902	EBV tethering prot
	45	70.8	1.1	2580	21	AAA75454	Nucleotide sequenc

ALIGNMENTS

```
RESULT 1
AAV02302
ID   AAV02302 standard; cDNA; 5200 BP.
XX
AC   AAV02302;
XX
DT   20-JUL-1998  (first entry)
XX
DE   Rat tumour suppressor gene SSeCKS.
XX
KW   SSeCKS; tumour suppressor gene; rat; protein kinase C; mitosis;
KW   cancer; malignancy; cell proliferation; Alzheimer's disease;
KW   therapy; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..4791
FT                   /*tag=  a
XX
PN   WO9740059-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;   97WO-US06830.
XX
PR   18-JUN-1996;   96US-0665401.
PR   19-APR-1996;   96US-0635121.
XX
PA   (GELM/) GELMAN I.
PA   (JAKE/) JAKEN S.
XX
```

```
PI   Gelman I,  Jaken S;
XX
DR   WPI; 1997-535770/49.
DR   P-PSDB; AAW31347.
XX
PT   Tumour suppressor gene SSeCKS - used as a mitotic regulator, and
PT   inhibitor of malignant phenotype
XX
PS   Claim 1; Fig 11A-L; 162pp; English.
XX
CC   This rat cDNA sequence codes for the full-length form of the novel
CC   tumour suppressor protein SSeCKs (see AAW31347).  The SSeCKs gene
CC   product is a substrate of protein kinase C and acts as a negative
CC   regulator of mitosis and as an inhibitor of the transformed
CC   phenotype.  A partial SSeCKS clone (see AAV02301) was identified by
CC   searching for cDNAs whose abundance was low in NIH 3T3 cells and
CC   decreased following the expression of the activated oncogene v-src.
CC   A 5'RACE product was spliced to this truncated sequence to construct
CC   the full-length SSeCKS cDNA sequence.  This full-length SSeCKS
CC   nucleic acid sequence, as well as homologous and hybridising nucleic
CC   acids are claimed, as are vectors comprising such nucleic acids,
CC   encoded proteins, host cells and methods of inhibiting the
CC   expression of a transformed phenotype in a host cell by introducing
CC   the nucleic acid.  Introduction of a SSeCKS nucleic acid or gene
CC   product into a host cell inhibits mitosis of the host cell,
CC   allowing the treatment of diseases associated with disorders of
CC   proliferation and/or with the expression of a malignant phenotype.
CC   SSeCKS can also be used to treat or identify disorders of
CC   cytoskeletal structure and cellular architecture (such as
CC   Alzheimer's disease), and may be a marker for aberrancies in
CC   fertility and/or nervous system development.
XX
SQ   Sequence 5200 BP; 1612 A; 1201 C; 1563 G; 824 T; 0 other;

  Query Match             84.1%;  Score 5182.4;  DB 18;  Length 5200;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 5189;  Conservative    0;  Mismatches   11;  Indels    0;  Gaps    0;

Qy         33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 60

Qy         93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 120

Qy        153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 180

Qy        213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 240

Qy        273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 300

Qy        333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 360

Qy        393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 420

Qy        453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 480

Qy        513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572
              ||||||||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GTTGGTTTTATTTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 540

Qy        573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 600

Qy        633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 660

Qy        693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 720

Qy        753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 780

Qy        813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 840

Qy        873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932
              |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 ACATCTTCCTCCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 900

Qy        933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 960

Qy        993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1020

Qy       1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1080

Qy       1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1140

Qy       1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1200

Qy       1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1260

Qy       1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1320

Qy       1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1380

Qy       1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1440

Qy       1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1500

Qy       1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1560

Qy       1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1620
```

```
Qy   1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1680

Qy   1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1740

Qy   1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1800

Qy   1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1860

Qy   1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1920

Qy   1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 1980

Qy   2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2040

Qy   2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2100

Qy   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2160

Qy   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252
          ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db   2161 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCTTTCCGAAGGG 2220

Qy   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2280

Qy   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2340

Qy   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2400

Qy   2433 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2460

Qy   2493 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2552
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2520

Qy   2553 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2580

Qy   2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2640

Qy   2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2700

Qy   2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2760

Qy   2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2820

Qy   2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2880

Qy   2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2940

Qy   2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3000

Qy   3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3060

Qy   3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3120

Qy   3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3180

Qy   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3240

Qy   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3300

Qy   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3360

Qy   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3420

Qy   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3480

Qy   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3540

Qy   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3600

Qy   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3660

Qy   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3720

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3780

Qy   3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db    3781 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3840

Qy    3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3841 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3900

Qy    3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3901 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3960

Qy    3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3961 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4020

Qy    4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4021 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4080

Qy    4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4081 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4140

Qy    4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4141 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4200

Qy    4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292
           |||||||  |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4201 GTAGGTGACCACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4260

Qy    4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4261 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4320

Qy    4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4321 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4380

Qy    4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4381 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4440

Qy    4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4441 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4500

Qy    4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4501 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4560

Qy    4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4652
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4561 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4620

Qy    4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4712
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4621 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4680

Qy    4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4681 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4740

Qy    4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGA 4832
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4741 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGA 4800

Qy    4833 AGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAA 4892
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4801 AGCTGCCTCCTCAACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAA 4860

Qy    4893 AGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACC 4952
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4861 AGCTTAGCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACC 4920

Qy    4953 AACGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAA 5012
           | |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db    4921 ACCGGACCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTCAGGTCCAGGAAGAAGAA 4980

Qy    5013 ATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCA 5072
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4981 ATGTCTACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCA 5040

Qy    5073 AAGGGAGACCTGGCAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGA 5132
           ||||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db    5041 AAGGGAGACCTGGCAGAATCCTCCGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGA 5100

Qy    5133 ATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAG 5192
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    5101 ATGTGAAAACAAGTCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAG 5160

Qy    5193 AGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTC 5232
           ||||||||| ||||||||||||||||||||||||||||||
Db    5161 AGCCCATGACATCCAGAGAGCAGGGCCGTCCAATGATTTC 5200


RESULT 2
AAV02301
ID   AAV02301 standard; cDNA; 5074 BP.
XX
AC   AAV02301;
XX
DT   20-JUL-1998  (first entry)
XX
DE   Rat tumour suppressor gene SSeCKS (truncated form).
XX
KW   SSeCKS; tumour suppressor gene; rat; protein kinase C; mitosis;
KW   cancer; malignancy; cell proliferation; Alzheimer's disease;
KW   therapy; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             176..4216
FT                   /*tag=  a
FT                   /transl_except= (pos:389..391, aa:Glu)
FT                   /transl_except= (pos:1661..1663, aa:Lys)
FT                   /transl_except= (pos:1668..1670, aa:Lys)
FT                   /transl_except= (pos:2444..2446, aa:Arg)
FT                   /transl_except= (pos:2528..2530, aa:Glu)
FT   polyA_signal    2528..2530
FT                   /*tag=  b
FT   polyA_signal    5056..5061
FT                   /*tag=  c
XX
PN   WO9740059-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;   97WO-US06830.
XX
PR   18-JUN-1996;   96US-0665401.
PR   19-APR-1996;   96US-0635121.
XX
PA   (GELM/) GELMAN I.
PA   (JAKE/) JAKEN S.
XX
PI   Gelman I,  Jaken S;
XX
DR   WPI; 1997-535770/49.
DR   P-PSDB; AAW31346.
XX
PT   Tumour suppressor gene SSeCKS - used as a mitotic regulator, and
PT   inhibitor of malignant phenotype
XX
PS   Example 6; Fig 3A-H; 162pp; English.
XX
```

```
CC   This rat cDNA sequence codes for an active truncated form of the
CC   novel tumour suppressor protein SSeCKs (see AAW31346).  The SSeCKs
CC   gene product is a substrate of protein kinase C and acts as a
CC   negative regulator of mitosis and as an inhibitor of the transformed
CC   phenotype.  The SSeCKS clone was identified by searching for cDNAs
CC   whose abundance was low in NIH 3T3 cells and decreased following the
CC   expression of the activated oncogene v-src.  A 5'RACE product was
CC   spliced to the truncated sequence to construct a full-length
CC   SSeCKS cDNA (see AAV02302).  This full-length SSeCKS nucleic acid
CC   sequence, as well as homologous and hybridising nucleic acids, are
CC   claimed, as are vectors comprising such nucleic acids, host cells
CC   and methods of inhibiting the expression of a transformed phenotype
CC   in a host cell by introducing such nucleic acids.  Introduction of
CC   a SSeCKS nucleic acid or gene product into a host cell inhibits
CC   mitosis of the host cell, allowing the treatment of diseases
CC   associated with disorders of proliferation and/or with the
CC   expression of a malignant phenotype.  SSeCKS can also be used to
CC   treat or identify disorders of cytoskeletal structure and cellular
CC   architecture (such as Alzheimer's disease), and may be a marker
CC   for aberrancies in fertility and/or nervous system development.
XX
SQ   Sequence 5074 BP; 1501 A; 1151 C; 1441 G; 981 T; 0 other;

  Query Match             76.2%;  Score 4696.6;  DB 18;  Length 5074;
  Best Local Similarity   96.7%;  Pred. No. 0;
  Matches 4985;  Conservative    0;  Mismatches   79;  Indels   91;  Gaps   15;

Qy       1016 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 1075
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 60

Qy       1076 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 1135
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 120

Qy       1136 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 1195
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 180

Qy       1196 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 1255
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 240

Qy       1256 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 1315
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 300

Qy       1316 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 1375
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 360

Qy       1376 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 1435
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 420

Qy       1436 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 1495
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 480

Qy       1496 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 1555
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 540

Qy       1556 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 1615
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 600

Qy       1616 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 1675
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 660

Qy       1676 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 1735
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 720

Qy       1736 AGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGA 1795
              |||||||||||||||| ||||||||||||||||||||||||||||||||||    |  ||
Db        721 AGGGCCGCTGGAAGCA-CCCAGGATGGGGAAGCTGAGGAAGGAACTACTTC----GTGGA 775

Qy       1796 GAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 1855
              ||  |  |||| | ||||||||||||||||||||||||||||||||||||||||||||||
Db        776 GAGAAGAAGAGGAAGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 835

Qy       1856 ACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 1915
              || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        836 AC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 894

Qy       1916 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 1975
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        895 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 954

Qy       1976 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 2035
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        955 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 1014

Qy       2036 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGA 2095
              ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db       1015 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCA---GATAT 1071

Qy       2096 AGGAGGGCCAAGGACACTGGGA-GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 2154
              | |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db       1072 AAGAGGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 1131

Qy       2155 AAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 2214
              |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1132 AAGAAGCC-GAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 1190

Qy       2215 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 2274
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1191 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 1250

Qy       2275 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCG 2334
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |
Db       1251 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGAAG 1310

Qy       2335 AAGACTCTAGTGTAGAGCA--------GTTGTCCACTGAGATCGAACCGAGTAGAGAAGA 2386
                |     | | |||            ||||||||||||||||||||| ||||||||||
Db       1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy       2387 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA 2446
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db       1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy       2447 -GCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTA 2505
               |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Db       1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy       2506 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAG 2565
              |||||||||| ||||||||||||||||||||| |||||||||||||||||||||||||||
Db       1491 ATGTCCCAGCAGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy       2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
              ||||||||||| |||||||||||        |||          ||||||||||||||||
Db       1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy       2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy       2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy       2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787
```

```
Qy   2806 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 2865
          ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTCAAGATGCCCATGACGACATGGTCA 1847

Qy   2866 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 2925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy   2926 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 2985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy   2986 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 3045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   3046 AGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 3105
          || |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   3106 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 3165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   3166 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 3225
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   3226 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 3285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   3286 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGG 3345
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   3346 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 3405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   3406 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 3465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   3466 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 3525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   3526 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 3585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   3586 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 3645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   3646 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCG 3705
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   3706 TTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 3765
          |   |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   3766 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 3825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   3826 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 3885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   3886 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 3945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   3946 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 4005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   4006 AAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 4065
          || | | |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   4066 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 4125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   4126 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 4185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   4186 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 4245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   4246 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 4305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   4306 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 4365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   4366 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 4425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   4426 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 4485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   4486 CAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 4545
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   4546 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 4605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   4606 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGG 4665
          |||||||||||||||||||||||||||||||||||||||||    |  | ||||||||||
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   4666 AGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCT 4785
          |||||||||||||||||||||||||||||||||||||||||       ||||||||||||
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGG--CTCAGCTCGGAAATGCT 3765

Qy   4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 3825

Qy   4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   3826 ACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 3884

Qy   4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
          |||||||    |  ||||||||||||||||||||||||||||||||||||||||||||||
Db   3885 CAGAGGC---AGTGTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 3941

Qy   4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
```

```
                ||||||||||| |     ||||||||||||||||||||||||||||||||||||||||||||||
Db     3942 TAACCGAGGAGCG-ATGCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 4000

Qy     5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 5085
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4001 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 4060

Qy     5086 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 5145
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4061 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 4120

Qy     5146 TCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATC 5205
            |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db     4121 TCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCATGAGATC 4179

Qy     5206 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 5265
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4180 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 4239

Qy     5266 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5325
            ||||||||||||||||||||
Db     4240 TCATCGGGAGCTAGAGCCAG---------------------------------------- 4259

Qy     5326 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5385
                                ||||||||||||||||||||||||||||||||||||||||
Db     4260 --------------------CTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4299

Qy     5386 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 5445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4300 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 4359

Qy     5446 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 5505
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4360 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 4419

Qy     5506 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 5565
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4420 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 4479

Qy     5566 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 5625
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4480 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 4539

Qy     5626 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 5685
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4540 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 4599

Qy     5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 5745
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4600 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 4659

Qy     5746 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 5805
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4660 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 4719

Qy     5806 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 5865
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4720 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 4779

Qy     5866 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 5925
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4780 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 4839

Qy     5926 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 5985
            ||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db     4840 GATGTTTCCTAATGCAGCAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 4899

Qy     5986 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 6045
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4900 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 4959

Qy     6046 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 6105
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4960 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 5019

Qy     6106 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5020 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 5074
```

```
RESULT 3
ABN59861
ID   ABN59861 standard; cDNA; 6614 BP.
XX
AC   ABN59861;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Novel human coding sequence SEQ ID NO: 272.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
XX
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97448.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Claim 1; SEQ ID NO 272; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention.
XX
SQ   Sequence 6614 BP; 2094 A; 1420 C; 1827 G; 1273 T; 0 other;

  Query Match             35.4%;  Score 2179.6;  DB 24;  Length 6614;
  Best Local Similarity   64.4%;  Pred. No. 0;
  Matches 4223;  Conservative    0;  Mismatches 1794;  Indels  537;  Gaps   43;

Qy        3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
            |||  |  |  | ||||| ||     ||||||||||| || || |||||||||||||
Db      162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy       63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
            |||| |||||||||    | |||||||  |||||||    ||||  | ||  ||| |||
Db      222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy      120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
            || |||  | || || ||  ||  |          || |     | |  |||  |  | |||||
```

```
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy    269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACAGAGGTTGGACAGAGA 521

Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    582 GTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCT 641

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    642 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 701

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    702 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 761

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    762 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 821

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    822 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 866

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    867 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 926

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    927 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCA---GTGGAGGAA 983

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    984 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1043

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1044 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1103

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1104 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1163

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1164 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1223

Qy   1029 CCAGCCTCGGAGGA---------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1224 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1283

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1284 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1343

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1344 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1403

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1404 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1463

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1464 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1499

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1500 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAA 1559

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1560 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1619

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1620 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1679

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1680 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1739

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1740 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1799

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1800 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1856

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1857 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1916

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1917 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1976

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1977 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2036

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2037 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2096

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2097 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2156

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2157 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2216

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2217 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2276

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2277 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2336

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2337 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2396

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2397 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2456
```

```
Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2457 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2516

Qy   2319 CTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2517 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2576

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2577 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2636

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2637 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2696

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2697 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2756

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2757 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2816

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2817 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2876

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2877 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2936

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2937 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2996

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2997 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3056

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3057 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3116

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3117 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3176

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3177 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3236

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3237 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3296

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3297 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3356

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3357 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3416

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3417 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3476

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3477 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3536

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3537 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3596

Qy   3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3597 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3656

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3657 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3716

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3717 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3776

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3777 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3836

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3837 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3896

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3897 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3956

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3957 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4016

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4017 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4076

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4077 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4136

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db   4137 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4196

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4197 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4256

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4257 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4316

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4317 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4376

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4377 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4436

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4437 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4496

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4497 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4556

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4557 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4616

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
```

```
Db   4617 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4676

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4677 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4736

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4737 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGGAGGAT 4790

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4791 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4850

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4851 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4910

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4911 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4970

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4971 GAAACGGAGAAAGAAGGAGAGGACCCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5030

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5031 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5090

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5091 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5150

Qy   4819 ------------------------------------------------------------ 4818
Db   5151 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5210

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5211 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5270

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5271 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5330

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5331 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5390

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5391 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5450

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5451 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5510

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5511 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5570

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5571 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5629

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5630 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5688

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5689 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5747

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5748 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5787

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5788 -------------------------------------------CATATAAATTCTGATTTA 5805

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5806 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5865

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5866 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5922

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5923 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5982

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5983 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6042

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6043 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6102

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6103 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6162

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6163 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6222

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6223 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6275

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6276 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6335

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6336 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6395

Qy   5954 TGTACGTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6396 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6453

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6454 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6513

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6514 TGCCATTCTTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6573

Qy   6134 CTGATTTAAATAAA 6147
Db   6574 CTGATTTAAATAAA 6587
```

RESULT 4
ABN96942
ID   ABN96942 standard; DNA; 6608 BP.
XX
AC   ABN96942;
XX
DT   13-AUG-2002 (first entry)

```
XX
DE   Gene #3440 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample  -
XX
PS   Claim 1; SEQ ID NO 3440; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6608 BP; 2091 A; 1418 C; 1827 G; 1272 T; 0 other;

  Query Match             34.8%;  Score 2144.4;  DB 24;  Length 6608;
  Best Local Similarity   64.3%;  Pred. No. 0;
  Matches 4215;  Conservative    0;  Mismatches 1796;  Indels  543;  Gaps   45;

Qy          3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db        162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy         63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db        222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy        120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db        282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy        171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db        342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy        231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db        402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy        269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db        462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy        309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db        522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy        369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db        582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy        429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db        639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy        489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db        699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy        549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db        759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy        609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db        819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy        669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db        864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy        729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db        924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy        789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db        981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy        849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db       1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy        909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db       1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy        969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db       1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy       1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db       1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy       1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db       1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy       1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db       1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy       1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db       1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy       1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db       1461 GAGGAGAGAACCGAAGAGCAGAAA------------------------ACGGAGGTGGAA 1496

Qy       1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
```

```
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713
```

```
Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3833

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3834 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3893

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3894 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3953

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3954 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4013

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4014 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4073

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4074 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4133

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4134 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4193

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4194 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4253

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4254 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4313

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4314 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4373

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4374 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4433

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4434 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4493

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4494 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4553

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4554 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4613

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4614 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4673

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4674 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4733

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4734 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4784

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4785 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4844

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4845 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4904

Qy   4617 GCTTATGATTCACAGACCCAGGTTC---------------CTGCATGCAGGCTTGACAG 4660
Db   4905 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4964

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4965 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5024

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5025 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5084

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5085 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5144

Qy   4819 ------------------------------------------------------------ 4818
Db   5145 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5204

Qy   4819 ------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5205 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5264

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5265 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5324

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5325 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5384

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5385 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5444

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5445 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5504

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5505 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5564

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5565 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5623

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5624 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5682

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5683 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5741

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5742 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5781

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5782 ------------------------------------------CATATAAATTCTGATTTA 5799

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
```

```
Db   5800 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5859

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
          |||| ||| | |||||  |||||||||||||||||| ||||    ||||||||| ||||
Db   5860 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5916

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
          |||||||||||||| |   ||                  |  ||||||||||||||| |
Db   5917 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5976

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
            || ||||| | |||      |||||| |   |  |||    ||||  |||||||   |
Db   5977 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6036

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
          ||      ||  ||| | | | | |                |   | ||  |    || ||
Db   6037 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6096

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
          |||| || ||| ||| ||                | | || ||||||| ||||   ||||||
Db   6097 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6156

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
          ||     |      |    |||   |||   ||||||||||||| ||| |||| | | | |
Db   6157 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6216

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
          |    ||  |||| |         || |||| |||||  ||||| |||||||||      |
Db   6217 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6269

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
           ||  |                  ||   |    | |   |   |   | |||||  || |||
Db   6270 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6329

Qy   5910 GTGCCTAAGGCCATGAGA---------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
          |||| |||  |   |                  |  |    | ||| ||||     | |||
Db   6330 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTATATGTTATGAAGAAAAGAATTGT 6389

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
          |||| |  |||||||| ||||||| |||||||||| | ||||| |       |    ||| |
Db   6390 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6447

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
          ||||  ||||||||| ||||||||||||||||||||| | ||| ||||  |||| ||
Db   6448 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6507

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
          ||||  | ||||||||||||||||||   |   |||  ||||||||| ||||||||||| |
Db   6508 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6567

Qy   6134 CTGATTTAAATAAA 6147
          ||||||||||||||
Db   6568 CTGATTTAAATAAA 6581
```

```
RESULT 5
ABL61938
ID   ABL61938 standard; DNA; 6608 BP.
XX
AC   ABL61938;
XX
DT   15-MAY-2002  (first entry)
XX
DE   Colon adenocarcinoma related gene sequence SEQ ID NO:275.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE,  Augustus M,  Carter KC,  Ebner R,  Endress G,  Horrigan S;
PI   Soppet DR,  Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set  -
XX
PS   Claim 1; SEQ ID 275; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
```

```
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 6608 BP; 2091 A; 1418 C; 1827 G; 1272 T; 0 other;

  Query Match             34.8%;  Score 2144.4;  DB 24;  Length 6608;
  Best Local Similarity   64.3%;  Pred. No. 0;
  Matches 4215;  Conservative    0;  Mismatches 1796;  Indels  543;  Gaps   45;

Qy          3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db        162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy         63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db        222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy        120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db        282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy        171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db        342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy        231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db        402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy        269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db        462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy        309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db        522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy        369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db        582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy        429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db        639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy        489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db        699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy        549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db        759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy        609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db        819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy        669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db        864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy        729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db        924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy        789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db        981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy        849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db       1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy        909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db       1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy        969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db       1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy       1029 CCAGCCTCGGAGGA-------------------------GCAGGAGCCGGCAGAAGAC 1061
Db       1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy       1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db       1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy       1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db       1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy       1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db       1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy       1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db       1461 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1496

Qy       1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db       1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy       1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db       1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy       1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db       1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy       1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db       1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy       1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db       1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy       1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db       1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy       1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db       1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy       1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db       1914 ATCACGTGTCTGGAAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy       1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db       1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy       1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db       2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy       1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
```

```
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3833

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3834 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3893

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3894 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3953

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3954 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4013

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4014 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4073

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4074 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4133

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db   4134 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4193

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4194 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4253

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4254 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4313
```

```
Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4314 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4373

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4374 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4433

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4434 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4493

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4494 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4553

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4554 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4613

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4614 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4673

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4674 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4733

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4734 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4784

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4785 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4844

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4845 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4904

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4905 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4964

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4965 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5024

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5025 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5084

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5085 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5144

Qy   4819 ------------------------------------------------------------ 4818
Db   5145 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5204

Qy   4819 -------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5205 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5264

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5265 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5324

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5325 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5384

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5385 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5444

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5445 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5504

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5505 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5564

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5565 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5623

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5624 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5682

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5683 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5741

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5742 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5781

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5782 ------------------------------------------CATATAAATTCTGATTTA 5799

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5800 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5859

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5860 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5916

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5917 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5976

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5977 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6036

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6037 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6096

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6097 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6156

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6157 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6216

Qy   5807 ACCATGATTTTTTCCCCCCTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6217 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6269

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6270 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6329

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6330 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6389

Qy   5954 TGTACGTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
```

```
Db    6390 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6447

Qy    6014 AGTCCTGTTCTTTACAGATGGTATTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073

Db    6448 AGTCAGGTTCTTTACAAATGGTATTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6507

Qy    6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133

Db    6508 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6567

Qy    6134 CTGATTTAAATAAA 6147

Db    6568 CTGATTTAAATAAA 6581
```

```
RESULT 6
AAV23545
ID   AAV23545 standard; cDNA; 6605 BP.
XX
AC   AAV23545;
XX
DT   13-JUL-1998  (first entry)
XX
DE   Human gravin polypeptide coding sequence.
XX
KW   Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW   cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW   Myasthenia gravis; nicotinic acetylcholine receptor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             192..5534
FT                   /*tag=  a
XX
PN   US5741890-A.
XX
PD   21-APR-1998.
XX
PF   19-DEC-1996;   96US-0769309.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Nauert JB,  Scott JD;
XX
DR   WPI; 1998-260552/23.
DR   P-PSDB; AAW53863.
XX
PT   New polypeptide fragments of protein kinase binding protein gravin -
PT   are useful for the study of modulation of action between gravin and
PT   protein kinase(s)
XX
PS   Example 1; Column 19-32; 32pp; English.
XX
CC   This sequence encodes the human gravin polypeptide, which represents
CC   a polypeptide of the invention. The polypeptides are fragments capable of
CC   binding to type II regulatory subunit of cAMP-dependent protein kinase
CC   (PKA). Gravin is a kinase anchoring protein that binds to type II
CC   regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC   antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC   develops antibodies against their own nicotinic acetylcholine receptors.
CC   The polypeptides are useful for providing analogues of gravin in the
CC   study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC   interactions between gravin and kinase. The peptides are involved in the
CC   modulation of gravin-kinase interactions.
XX
SQ   Sequence 6605 BP; 2091 A; 1416 C; 1826 G; 1272 T; 0 other;

  Query Match             34.4%;  Score 2118.8;  DB 19;  Length 6605;
  Best Local Similarity   64.2%;  Pred. No. 0;
  Matches 4206;  Conservative    0;  Mismatches 1802;  Indels  546;  Gaps   46;
```

```
Qy       3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62

Db     162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy      63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119

Db     222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy     120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170

Db     282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy     171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230

Db     342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy     231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268

Db     402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy     269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308

Db     462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy     309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368

Db     522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy     369 GTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428

Db     582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy     429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488

Db     639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy     489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548

Db     699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy     549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608

Db     759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy     609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668

Db     819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy     669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728

Db     864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy     729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788

Db     924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy     789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848

Db     981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy     849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908

Db    1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy     909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968

Db    1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy     969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028

Db    1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220
```

```
Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1461 GAGGAGAGAACCGAAGAGCAGAAA------------------------ACGGAGGTGGAA 1496

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
```

```
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGA---GGTGCATCTGGTACCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCT 3830

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3831 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3890

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3891 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3950

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3951 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4010

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4011 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4070

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4071 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4130

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4131 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4190

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4191 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4250

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4251 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4310

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4311 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4370

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4371 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4430

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4431 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4490

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4491 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4550

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4551 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4610

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4611 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4670

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4671 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4730

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4731 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4781

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4782 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4841

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4842 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4901

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4902 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4961

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4962 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5021

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5022 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5081

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5082 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5141

Qy   4819 ------------------------------------------------------------ 4818
Db   5142 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5201

Qy   4819 -------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5202 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5261

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5262 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5321

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5322 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5381

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5382 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5441

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5442 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5501

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5502 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5561
```

```
Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5562 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5620

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5621 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5679

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5680 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5738

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5739 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5778

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5779 ------------------------------------------CATATAAATTCTGATTTA 5796

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5797 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5856

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5857 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5913

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5914 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5973

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5974 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6033

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6034 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6093

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6094 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6153

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6154 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6213

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6214 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6266

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6267 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6326

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6327 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6386

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6387 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6444

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6445 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6504

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6505 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6564

Qy   6134 CTGATTTAAATAAA 6147
Db   6565 CTGATTTAAATAAA 6578
```

RESULT 7
AAA74903

```
ID   AAA74903 standard; cDNA; 6605 BP.
XX
AC   AAA74903;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Human gravin coding sequence.
XX
KW   Human; gravin; PKA RII binding site; myasthenia gravis;
KW   kinase anchoring protein; cAMP dependent protein kinase; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             192..5534
FT                   /*tag=  a
FT                   /product= "gravin"
XX
PN   US6090929-A.
XX
PD   18-JUL-2000.
XX
PF   19-DEC-1997;   97US-0994570.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Scott JD,  Nauert JB;
XX
DR   WPI; 2000-523763/47.
DR   P-PSDB; AAB15380.
XX
PT   Novel polynucleotides useful for detecting gravin in patients suffering
PT   from Myasthenia gravis encodes cAMP-dependent protein kinase-binding
PT   polypeptide and protein kinase C-binding polypeptide of gravin  -
XX
PS   Example 1; Column 23-36; 34pp; English.
XX
CC   The present sequence is the coding sequence for human gravin. Gravin is
CC   an A-kinase anchoring protein (AKAP) which is involved in the
CC   localisation of cAMP dependent protein kinase A (PKA) via interactions
CC   between the RII binding region and the PKA regulatory subunit RII. Gravin
CC   is also an antigen found in myasthenia gravis sufferers, and it is
CC   thought that antibodies to it may be useful in modulating the binding of
CC   PKA, and thus aid in the treatment of the disease. This sequence was
CC   isolated by first screening a human umbilical vein endothelial cell cDNA
CC   library with serum from a myasthenia gravis patient, and then searching a
CC   human heart cDNA library for sequences resembling the isolated sequence.
CC   This was done because the first sequence obtained was shown to be shorter
CC   than the full length coding sequence.
XX
SQ   Sequence 6605 BP; 2091 A; 1416 C; 1826 G; 1272 T; 0 other;

  Query Match             34.4%;  Score 2118.8;  DB 21;  Length 6605;
  Best Local Similarity   64.2%;  Pred. No. 0;
  Matches 4206;  Conservative    0;  Mismatches 1802;  Indels  546;  Gaps   46;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
```

```
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341
Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401
Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG-------------------- 268
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461
Qy    269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521
Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db    522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581
Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638
Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698
Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758
Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818
Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863
Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923
Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980
Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040
Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100
Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160
Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220
Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280
Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340
Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400
Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460
Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1461 GAGGAGAGAACCGAAGAGCAGAAA------------------------ACGGAGGTGGAA 1496
Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556
Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616
Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676
Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736
Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796
Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853
Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913
Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973
Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033
Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093
Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153
Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213
Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273
Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333
Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393
Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453
```

```
Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGA---GGTGCATCTGGTACCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCT 3830

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3831 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3890

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3891 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3950

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3951 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4010

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4011 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4070

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4071 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4130

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4131 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4190

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4191 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4250

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4251 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4310

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4311 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4370

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4371 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4430

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4431 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4490

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4491 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4550

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4551 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4610

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
```

```
Db   4611 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4670

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4671 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4730

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4731 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4781

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4782 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4841

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4842 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4901

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4902 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4961

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4962 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5021

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5022 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5081

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5082 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5141

Qy   4819 ------------------------------------------------------------ 4818
Db   5142 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5201

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5202 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5261

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5262 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5321

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5322 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5381

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5382 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5441

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5442 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5501

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5502 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5561

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5562 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5620

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5621 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5679

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5680 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5738

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5739 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5778

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5779 ------------------------------------------CATATAAATTCTGATTTA 5796

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5797 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5856

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5857 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5913

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5914 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5973

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5974 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6033

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6034 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6093

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6094 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6153

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6154 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6213

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6214 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6266

Qy   5867 TGTCAGA----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6267 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6326

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6327 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6386

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6387 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6444

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6445 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6504

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6505 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6564

Qy   6134 CTGATTTAAATAAA 6147
Db   6565 CTGATTTAAATAAA 6578
```

RESULT 8
AAS85205
ID AAS85205 standard; cDNA; 6886 BP.
XX
AC AAS85205;
XX

```
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #21009.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG21018.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 21009; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6886 BP; 2142 A; 1483 C; 1908 G; 1352 T; 1 other;

  Query Match             31.2%;  Score 1920.6;  DB 23;  Length 6886;
  Best Local Similarity   67.3%;  Pred. No. 0;
  Matches 3255;  Conservative    0;  Mismatches 1334;  Indels  244;  Gaps   29;

Qy          3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db        162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy         63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db        222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy        120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db        282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy        171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db        342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy        231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268
Db        402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy        269 -------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db        462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCCGGCAGGTTGATCACATA 521

Qy        309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db        522 GACTCTGAAGATGTGAGCGACAGAGACTCCGATACAGAGATGACTACTAAGTCAGCGGTA 581

Qy        369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db        582 GTTCACGACATCACAGATGATGGGCAGGAGGAGACACCCGATATAATCGAACAGATTCCT 641

Qy        429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db        642 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 701

Qy        489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db        702 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 761

Qy        549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db        762 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 821

Qy        609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db        822 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 866

Qy        669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db        867 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 926

Qy        729 GAACAGAGCAGCACAGAAA-TCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGA 787
Db        927 GAGCAAAGCCACGCAGAAATTTCTCCCCCAAGCGGAATCTGGCCAA---GCAGTGGAGGA 983

Qy        788 AGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCC 847
Db        984 ATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCC 1043

Qy        848 GAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTG 907
Db       1044 GACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTG 1103

Qy        908 GGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGC 967
Db       1104 GGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTC 1163

Qy        968 CGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGA 1027
Db       1164 AGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGA 1223

Qy       1028 GCCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGA 1060
Db       1224 GGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAG 1283

Qy       1061 CACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCA 1120
Db       1284 TGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCA 1343

Qy       1121 GGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTT 1180
Db       1344 AGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTT 1403

Qy       1181 TGATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGT 1237
```

```
Db   1404 TGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGT 1463

Qy   1238 GGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGA 1297
Db   1464 GGAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGA 1499

Qy   1298 AGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGA 1357
Db   1500 AGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGA 1559

Qy   1358 AGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCA 1417
Db   1560 AGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCC 1619

Qy   1418 CACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCAT 1477
Db   1620 TACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGT 1679

Qy   1478 TGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTT 1537
Db   1680 TGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACT 1739

Qy   1538 GAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAA 1597
Db   1740 AAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAA 1799

Qy   1598 ACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCC 1657
Db   1800 AAGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCC 1856

Qy   1658 AGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGA 1717
Db   1857 GGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGA 1916

Qy   1718 GACCACGTGTCTGGAGAAAGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGG 1777
Db   1917 GATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGG 1976

Qy   1778 AACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAA 1837
Db   1977 AGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAA 2036

Qy   1838 GATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCT 1897
Db   2037 GATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCT 2096

Qy   1898 GGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGA 1957
Db   2097 GGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGA 2156

Qy   1958 TGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATAC 2017
Db   2157 AGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATAC 2216

Qy   2018 TTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGC 2077
Db   2217 CTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGG 2276

Qy   2078 ATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGA 2137
Db   2277 GTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGA 2336

Qy   2138 GGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCA 2197
Db   2337 TGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACA 2396

Qy   2198 GGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGG 2257
Db   2397 TGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGG 2456

Qy   2258 TGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAA 2317
Db   2457 CGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAGGAAAAAAATCAAAGTCCAA 2516

Qy   2318 ACTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGA 2365
Db   2517 GCTGGAAGAGAAAAGCGAAGACTCCCTAGCTGGGTCTGGTGTAGAACATTCCACTCCAGA 2576

Qy   2366 GATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCG 2425
Db   2577 CACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACTAAG 2636

Qy   2426 GAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGA 2485
Db   2637 GAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGG 2696

Qy   2486 GATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGT 2545
Db   2697 GGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGT 2756

Qy   2546 GGAGAGGGAGAAGATGGAAGCCCAG----------GGGAATACGGAGCTGCCCCAGCTGC 2595
Db   2757 AGAAAGGGAGAAAATGGAGGCCCAGCAAGCCCAAAAAAAGCGCAGAGCAGCCCGAGCAGA 2816

Qy   2596 TGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCG 2655
Db   2817 AGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAG 2876

Qy   2656 CAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATAT 2715
Db   2877 CTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATAT 2936

Qy   2716 CCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAG 2775
Db   2937 CTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGG 2996

Qy   2776 AGGTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTAC 2832
Db   2997 AGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGC 3056

Qy   2833 CAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAG 2892
Db   3057 CAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAG 3116

Qy   2893 CTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGG 2952
Db   3117 CTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTG 3176

Qy   2953 GGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACA 3012
Db   3177 CTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACA 3236

Qy   3013 CCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAG 3072
Db   3237 CCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGC 3296

Qy   3073 AGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCC 3132
Db   3297 AAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCC 3356

Qy   3133 AGGTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTG 3189
Db   3357 AGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAG 3416

Qy   3190 AGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGA 3249
Db   3417 AAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGA 3476

Qy   3250 AAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGG 3309
Db   3477 AAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGAAGGTGGTGGGGCAGA 3536

Qy   3310 CTACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTA---------GACC 3357
Db   3537 CCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGC 3596
```

```
Qy   3358 ATGTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGG 3399
Db   3597 TTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGTAAAATCACAGGAGATGGTGATGG 3656

Qy   3400 AACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCA 3459
Db   3657 AACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCA 3716

Qy   3460 CTCCCTTAGCAGATTCAGACACTGCAGATGGACACAGCAAGATGAAACCATTGACAGCC 3519
Db   3717 CCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCC 3776

Qy   3520 AGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTA 3579
Db   3777 ATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTC 3836

Qy   3580 CTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATG 3639
Db   3837 CTGCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAG 3896

Qy   3640 GGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCG 3699
Db   3897 AACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTG 3956

Qy   3700 TGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGT 3750
Db   3957 TATCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAA 4016

Qy   3751 TGGATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCA 3804
Db   4017 CCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCA 4076

Qy   3805 ACAGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTC 3864
Db   4077 GTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTA 4136

Qy   3865 AGGAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGG 3909
Db   4137 AAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGA 4196

Qy   3910 GAGAGATGGAAACTGACGTTGAAAAGGAGAAA---AGGGAGACAAAGCCAGAGCAAGTGA 3966
Db   4197 GAGAGATGGTAGTTCAAGTCGAAAGGGAGGAATACCAGTAGCAGAGCCCAACCCATGTGA 4256

Qy   3967 GTGAAG---AAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGC 4023
Db   4257 ATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGC 4316

Qy   4024 CAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAG 4083
Db   4317 TCCTCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAG 4376

Qy   4084 GAAGCCCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
Db   4377 GAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGA 4436

Qy   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCAT------AGAAA 4194
Db   4437 GCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAA 4496

Qy   4195 CGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAG 4254
Db   4497 CTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGG 4556

Qy   4255 CTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTAC 4314
Db   4557 AAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGC 4616

Qy   4315 CCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAA 4374
Db   4617 CCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGA 4676

Qy   4375 GCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGG 4434
Db   4677 AAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATG 4736

Qy   4435 AAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAA 4494
Db   4737 AAGTCGATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTGAGG 4790

Qy   4495 AAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGA-TTGTG 4553
Db   4791 AGGATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAAACTTGTC 4850

Qy   4554 CTGAACGTCATTCAGACAGCCGTTGACCAG-TTCGCACGTAC---AGAAACAGCCCCCGA 4609
Db   4851 CAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGA 4910

Qy   4610 AACTCATGCTTATGATTCACAGACCCAGGTTCC 4642
Db   4911 AATGTTGACGTCTGAGTTACAGACACAAGCTTC 4943
```

RESULT 9
ABK63316/c
ID   ABK63316 standard; cDNA; 563 BP.
XX
AC   ABK63316;
XX
DT   18-JUN-2002 (first entry)
XX
DE   Rat sequence differentially expressed in response to a hepatotoxin #1223.
XX
KW   Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KW   differential expression; centrilobular necrosis; steatosis.
XX
OS   Rattus norvegicus.
XX
PN   WO200210453-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-US23872.
XX
PR   31-JUL-2000; 2000US-222040P.
PR   02-NOV-2000; 2000US-244880P.
PR   11-MAY-2001; 2001US-290029P.
PR   15-MAY-2001; 2001US-290645P.
PR   22-MAY-2001; 2001US-292336P.
PR   06-JUN-2001; 2001US-295798P.
PR   13-JUN-2001; 2001US-297457P.
PR   19-JUN-2001; 2001US-298884P.
PR   09-JUL-2001; 2001US-303459P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick D, Porter MW, Johnson KR, Castle AL, Elashoff MR;
XX
DR   WPI; 2002-241625/29.
XX
PT   Predicting toxic effects of compounds or the progression of these toxic
PT   effects by determining the changes in gene expression in tissues or
PT   cells exposed to the toxin and comparing these to gene expression in
PT   unexposed tissues or cells -
XX
PS   Claim 1; Seq ID No 1223; 239pp; English.
XX
CC   The invention relates to methods for predicting toxic effects of
CC   compounds or the progression of these toxic effects by determining the
CC   global changes in gene expression in tissues or cells exposed to the
CC   toxin and comparing these to gene expression in unexposed tissues or
CC   cells. Also included are methods of predicting at least one toxic
CC   effect of a compound or progression of a toxic effect, preferably the
CC   hepatotoxicity of a compound, comprising detecting the level of
CC   expression in a tissue or cell sample exposed to the compound of two or
CC   more genes listed in the specification, where differential expression of

```
CC   the genes is indicative of at least one toxic effect or progression.
CC   The method can also be used to identify an agent which modulates the
CC   toxic response and predict cellular pathways that a compound modulates
CC   in a cell. The methods utilise a set of at least two probes (on a solid
CC   support in kit form), where each of the probes comprises a sequence that
CC   specifically hybridises to a gene listed in the specification, a computer
CC   system comprising a database containing information identifying the
CC   expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC   set of genes comprising at least two genes listed in the specification,
CC   and a user interface to view the information used to present information
CC   identifying the expression level in a tissue or cell of at least one gene
CC   listed in the specification. The method is useful for elucidating global
CC   changes in gene expression and for identifying toxicity markers in
CC   tissues or cell exposed to a known toxin. The genes may be used as
CC   toxicity markers in drug screening and toxicity assays. The genes and
CC   gene expression information may be used as diagnostic markers for the
CC   prediction or identification of the physiological state of tissue or cell
CC   sample that has been exposed to a compound or agent. Hepatotoxicity
CC   is characterised by centrilobular necrosis and steatosis. The present
CC   sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC   which is differentially expressed in response to a hepatotoxic agent.
XX
SQ   Sequence 563 BP; 191 A; 107 C; 110 G; 155 T; 0 other;

  Query Match              8.9%;  Score 550;  DB 24;  Length 563;
  Best Local Similarity   99.1%;  Pred. No. 8e-132;
  Matches  553;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5601 ATAGCGACGCTTTGAAAAGCTCCAAGCTCCAACTGTAACCTGCAGCAAACAGATAACATT 5660
              | |||||||||||||||||||||||||| |||||||||||||||||||||| ||||||||
Db        563 AAAGCGACGCTTTGAAAAGCTCCAAGCTTCAACTGTAACCTGCAGCAAACAAATAACATT 504

Qy       5661 CCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCT 5720
              ||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db        503 CCTGGCAAGAAAAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCT 444

Qy       5721 AGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTT 5780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        443 AGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTT 384

Qy       5781 TGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAG 5840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        383 TGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAG 324

Qy       5841 TCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGA 5900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        323 TCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGA 264

Qy       5901 CAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGT 5960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        263 CAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGT 204

Qy       5961 TTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTG 6020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        203 TTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTG 144

Qy       6021 TTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCT 6080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        143 TTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCT 84

Qy       6081 TCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 6140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         83 TCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 24

Qy       6141 AAATAAACATATTTGACT 6158
              ||||||||||||||| ||
Db         23 AAATAAACATATTTGCCT 6


RESULT 10
AAX40328
ID   AAX40328 standard; cDNA; 377 BP.
XX
AC   AAX40328;
XX
DT   16-JUN-1999  (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:115.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906439-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998;   98WO-IB01233.
XX
PR   01-AUG-1997;   97US-0904468.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A,  Dumas Milne Edwards J,  Lacroix B;
XX
DR   WPI; 1999-153700/13.
DR   P-PSDB; AAY11610.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
PT   cDNA libraries derived from liver, lung, large intestine, colon,
PT   thyroid and pancreas tissue
XX
PS   Claim 1; Page 228-229; 398pp; English.
XX
CC   AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for
CC   human secreted proteins, and encode the proteins given in AAY11533 to
CC   AAY11679, respectively. The proteins given represent the signal peptide
CC   and an N-terminal fragment of a secreted protein. The nucleic acid
CC   sequences can be used for producing secreted human gene products. They
CC   can also be used to develop products for diagnosis and therapy. The
CC   proteins obtained may have cytokine activity, cell
CC   proliferation/differentiation activity, haematopoiesis regulating
CC   activity, tissue growth regulating activity, reproductive hormone
CC   regulating activity, chemotactic/ chemokinetic activity, haemostatic and
CC   thrombolytic activity, receptor/ ligand activity, anti-inflammatory
CC   activity, tumour inhibition activity or other activities. The products
CC   can be used in forensic, gene therapy and chromosome mapping procedures.
CC   The sequences can also be used for obtaining corresponding promoter
CC   sequences. The nucleic acids encoding the signal peptide can be used for
CC   directing extracellular secretion of a polypeptide or the insertion of a
CC   polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ   Sequence 377 BP; 111 A; 85 C; 113 G; 66 T; 2 other;

  Query Match              2.8%;  Score 172.2;  DB 20;  Length 377;
  Best Local Similarity   68.9%;  Pred. No. 4.3e-34;
  Matches  253;  Conservative    1;  Mismatches  104;  Indels    9;  Gaps    1;

Qy       2455 AAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAG 2514
              ||||| |||| |||||| ||||||||   | |   || || || ||| || ||||||| |
Db          1 AAGCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGG 60

Qy       2515 CCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGG---------AAG 2565
              |||| || ||||||||||||| |||| || || |||||||| ||||          |||
Db         61 CCGTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAG 120

Qy       2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
              ||||    |   | |||| |||| ||| |  ||  ||       | ||||| :|||||||||
Db        121 CCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCKSAGGAGCTCA 180

Qy       2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
```

```
             | ||| || ||| || |    ||    |  || |||   ||| ||||| |||||||  ||||
Db      181 GCGAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCA 240

Qy     2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
              | | |||||  |||||||||| ||||||| ||||| || ||||||||||||||||
Db      241 TTATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAG 300

Qy     2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
              || |  ||||||   |          ||| |||||     |||| |||   | |||||||||
Db      301 TAGAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAG 360

Qy     2806 AAACTCC 2812
             ||   ||
Db      361 AAGAACC 367
```

```
RESULT 11
ABK63568
ID   ABK63568 standard; cDNA; 178 BP.
XX
AC   ABK63568;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Rat sequence differentially expressed in response to a hepatotoxin #1475.
XX
KW   Rat; ss; hepatotoxin; expressed sequence tag; EST; drug screening;
KW   differential expression; centrilobular necrosis; steatosis.
XX
OS   Rattus norvegicus.
XX
PN   WO200210453-A2.
XX
PD   07-FEB-2002.
XX
PF   30-JUL-2001; 2001WO-US23872.
XX
PR   31-JUL-2000; 2000US-222040P.
PR   02-NOV-2000; 2000US-244880P.
PR   11-MAY-2001; 2001US-290029P.
PR   15-MAY-2001; 2001US-290645P.
PR   22-MAY-2001; 2001US-292336P.
PR   06-JUN-2001; 2001US-295798P.
PR   13-JUN-2001; 2001US-297457P.
PR   19-JUN-2001; 2001US-298884P.
PR   09-JUL-2001; 2001US-303459P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Mendrick D,  Porter MW,  Johnson KR,  Castle AL,  Elashoff MR;
XX
DR   WPI; 2002-241625/29.
XX
PT   Predicting toxic effects of compounds or the progression of these toxic
PT   effects by determining the changes in gene expression in tissues or
PT   cells exposed to the toxin and comparing these to gene expression in
PT   unexposed tissues or cells -
XX
PS   Claim 1; Seq ID No 1475; 239pp; English.
XX
CC   The invention relates to methods for predicting toxic effects of
CC   compounds or the progression of these toxic effects by determining the
CC   global changes in gene expression in tissues or cells exposed to the
CC   toxin and comparing these to gene expression in unexposed tissues or
CC   cells. Also included are methods of predicting at least one toxic
CC   effect of a compound or progression of a toxic effect, preferably the
CC   hepatotoxicity of a compound, comprising detecting the level of
CC   expression in a tissue or cell sample exposed to the compound of two or
CC   more genes listed in the specification, where  differential expression of
CC   the genes is indicative of at least one toxic effect or progression.
CC   The method can also be used to identify an agent which modulates the
CC   toxic response and predict cellular pathways that a compound modulates
CC   in a cell. The methods utilise a set of at least two probes (on a solid
CC   support in kit form), where each of the probes comprises a sequence that
CC   specifically hybridises to a gene listed in the specification, a computer
CC   system comprising a database containing information identifying the
CC   expression level in a tissue or cell sample exposed to a hepatotoxin of a
CC   set of genes comprising at least two genes listed in the specification,
CC   and a user interface to view the information used to present information
CC   identifying the expression level in a tissue or cell of at least one gene
CC   listed in the specification. The method is useful for elucidating global
CC   changes in gene expression and for identifying toxicity markers in
CC   tissues or cell exposed to a known toxin. The genes may be used as
CC   toxicity markers in drug screening and toxicity assays. The genes and
CC   gene expression information may be used as diagnostic markers for the
CC   prediction or identification of the physiological state of tissue or cell
CC   sample that has been exposed to a compound or agent. Hepatotoxicity
CC   is characterised by centrilobular necrosis and steatosis. The present
CC   sequence is an expressed sequence tag (EST) or cDNA derived from a gene
CC   which is differentially expressed in response to a hepatotoxic agent.
XX
SQ   Sequence 178 BP; 41 A; 30 C; 33 G; 74 T; 0 other;

  Query Match             2.6%;  Score 162.2;  DB 24;  Length 178;
  Best Local Similarity   97.8%;  Pred. No. 1.1e-31;
  Matches  175;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy     5962 TTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTGT 6021
            |||||||||||||||||||||||||||||||||||||||||||||||||| | ||||||
Db        1 TTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTCACTCCTGT 60

Qy     6022 TCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTT 6081
            ||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||
Db       61 TCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGT-CCCCTT 119

Qy     6082 CTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 6140
            |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db      120 CTTTAAGAACAATGTTGCATTACGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 178


RESULT 12
AAV86620
ID   AAV86620 standard; cDNA; 483 BP.
XX
AC   AAV86620;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone BF146.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
```

```
PS   Claim 1; Page 294; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 483 BP; 156 A; 103 C; 133 G; 90 T; 1 other;

  Query Match             1.7%;  Score 105.2;  DB 20;  Length 483;
  Best Local Similarity   60.4%;  Pred. No. 1.1e-16;
  Matches  236;  Conservative    0;  Mismatches  134;  Indels   21;  Gaps    3;

Qy   3811 AAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAA 3870

Db     44 AAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAGG 103

Qy   3871 AGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGAGAGA 3915

Db    104 ATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAGA 163

Qy   3916 TGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAG 3975

Db    164 TGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAGA 223

Qy   3976 G---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGA 4032

Db    224 AGCTTGAGCACGAAACAGCTGTTACCGNATCTGAAGAGGTCAGTAAGCAGCTCCTCCAGA 283

Qy   4033 CACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTT 4092

Db    284 CAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCTC 343

Qy   4093 CTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACA 4149

Db    344 CTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAGG 403

Qy   4150 CAACAGTCACTCAAACAGCAGAAGCTGTGGA 4180

Db    404 CATCATTCACTCTAACAGCGGCTGCAGAGGA 434


RESULT 13
ABV56830/c
ID   ABV56830 standard; cDNA; 459 BP.
XX
AC   ABV56830;
XX
DT   17-SEP-2002  (first entry)
XX
DE   Human prostate expression marker cDNA 56821.
XX
KW   Human; prostate cancer; cytostatic; carcinogen; pharmacodyanamic marker;
KW   pharmacogenomic marker; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200160860-A2.
XX
PD   23-AUG-2001.
XX
PF   20-FEB-2001; 2001WO-US05171.
XX
PR   17-FEB-2000; 2000US-183319P.
PR   16-MAR-2000; 2000US-189862P.
PR   25-MAY-2000; 2000US-207454P.
PR   09-JUN-2000; 2000US-211314P.
PR   18-JUL-2000; 2000US-219007P.
PR   13-DEC-2000; 2000US-255281P.
XX
PA   (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.
XX
PI   Schlegel R,  Endege WO,  Monahan JE;
XX
DR   WPI; 2001-662795/76.
XX
PT   Novel isolated nucleic acid molecule associated with cancerous state of
PT   prostate cells and correlating with presence of prostate cancer, useful
PT   for detecting presence of prostate cancer, stage of prostate cancer  -
XX
PS   Claim 1; Page 10950; 11750pp; English.
XX
CC   The invention relates to an isolated nucleic acid molecule (I) comprising
CC   a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the
CC   specification or its complement. (I) is useful for:
CC   (a) assessing whether a patient is afflicted with prostate cancer;
CC   (b) monitoring the progression of prostate cancer in a patient;
CC   (c) assessing the efficacy of a test compound to inhibit prostate
CC   cancer in a patient;
CC   (d) assessing the efficacy of a therapy for inhibiting prostate cancer
CC   in a patient;
CC   (e) selecting a composition for inhibiting prostate cancer in a patient;
CC   (f) assessing the prostate cell carcinogenic potential of a compound;
CC   (g) determining whether prostate cancer has metastasized in a patient;
CC   (h) assessing the aggressiveness or indolence of prostate cancer in a
CC   patient;
CC   (I) is also useful as a pharmacodyanamic or pharmacogenomic marker.
XX
SQ   Sequence 459 BP; 143 A; 99 C; 59 G; 158 T; 0 other;

  Query Match             1.7%;  Score 101.8;  DB 23;  Length 459;
  Best Local Similarity   70.5%;  Pred. No. 8.3e-16;
  Matches  136;  Conservative    0;  Mismatches   57;  Indels    0;  Gaps    0;

Qy   5963 TTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTT 6022

Db    235 TTTTGATTTTACTTTTAAAAGGTGGGCGGCATTTACACAAGGGATGAAATAAGTCAGGTT 176

Qy   6023 CTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTC 6082

Db    175 TTTTACAAAAGGGATTTTGAAAAAAACTGGATTGGGTTTGGGCCAAATTTGGGCCCTTTT 116

Qy   6083 TTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTTAA 6142

Db    115 TTTAAGAACAAAGGTGCCACCCATTCATTTGGATAAGTTGTGATTTGACGACTGATTTAA 56

Qy   6143 ATAAACATATTTG 6155

Db     55 AAAAAAAATTTGG 43


RESULT 14
AAI86887
ID   AAI86887 standard; cDNA; 400 BP.
XX
AC   AAI86887;
XX
DT   06-NOV-2001  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 6947.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
```

```
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO06956.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 6947; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 400 BP; 138 A; 80 C; 90 G; 92 T; 0 other;

  Query Match             1.5%;  Score 95.4;  DB 22;  Length 400;
  Best Local Similarity   63.3%;  Pred. No. 3.5e-14;
  Matches  241;  Conservative    0;  Mismatches  111;  Indels   29;  Gaps    5;

Qy   4976 GGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGA 5035

Db      1 GGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCACAGTGAATCAGATAAAGC 60

Qy   5036 GAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTA 5095

Db     61 GATCACACCCCACTCACAGGAGGAGTTTCAGAAACAAGAGAGAGAATCTGCCGAGTGCTG 120

Qy   5096 AGATGTTAGTTG----------------------CTCATTGTACATCTGTAAGACCAGAA 5133

Db    121 AACTTACAGAATCTTAAAACATGGTGCAGTTAAACTCATTGGCTGTTTGGAAGACCAGAA 180

Qy   5134 TGTGAAAACAAGTCACAGAACAAG----ATGCTGCTGTTGGGACCTTGAGACCAAGATTT 5189

Db    181 TGTGAAGACCGGGAATAGAAGAAAATGAATGCTGCTGCTGATA-CTGAAGACCAACATTT 239

Qy   5190 CAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCC 5249

Db    240 CAGAACCT-TGAGAATTGGAGAGCAGGCACATCAACTGATATCATTTCTAGAGAGC-CCC 297

Qy   5250 CGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACT 5309

Db    298 TGACGATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATTTCCTCTTTTCAAGACC 357

Qy   5310 GCCTTTGATTTGCCCCTTGAT 5330

Db    358 AACCTACAATTTTTCCTTGAT 378
```

```
RESULT 15
AAV86330
ID   AAV86330 standard; cDNA; 339 BP.
XX
AC   AAV86330;
XX
DT   27-APR-1999  (first entry)
XX
DE   EST clone AJ169.
XX
KW   Expressed sequence tag; secreted protein; haematopoiesis regulator;
KW   tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW   chemotaxis; chemokinesis; haemostasis; gene therapy; thrombolysis;
KW   receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
XX
OS   Homo sapiens.
XX
PN   WO9845435-A2.
XX
PD   15-OCT-1998.
XX
PF   10-APR-1998;   98WO-US06954.
XX
PR   10-APR-1997;   97US-0835913.
XX
PA   (GEMY ) GENETICS INST INC.
XX
PI   Agostino MJ,  Jacobs K,  Lavallie ER,  McCoy JM,  Merberg D;
PI   Racie LA,  Spaulding V,  Treacy M;
XX
DR   WPI; 1999-070076/06.
XX
PT   New polynucleotides encoding human secreted proteins - derived from
PT   e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT   ovary, pituitary, retina and colon cDNA libraries
XX
PS   Claim 1; Page 198; 633pp; English.
XX
CC   This sequence represents an expressed sequence tag (EST), and is a
CC   polynucleotide of the invention. The polynucleotides of the invention are
CC   all secreted EST sequences isolated from a variety of human tissue
CC   sources. The EST sequences and proteins encoded by them are predicted to
CC   have useful biological activities which would make them suitable for
CC   treating, preventing or ameliorating medical conditions in humans and
CC   animals, although no supporting data is given. Suggested activities
CC   include nutritional activity, immune stimulating or suppressing activity,
CC   haematopoiesis regulating activity, tissue growth activity,
CC   activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC   and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC   activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC   activity. The EST sequences are also stated to be useful for gene
CC   therapy.
XX
SQ   Sequence 339 BP; 109 A; 74 C; 91 G; 65 T; 0 other;

  Query Match             1.5%;  Score 92;  DB 20;  Length 339;
  Best Local Similarity   58.3%;  Pred. No. 2.4e-13;
  Matches  161;  Conservative    0;  Mismatches  115;  Indels    0;  Gaps    0;

Qy   3464 CTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGA 3523

Db     12 CATGGCCTACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGA 71

Qy   3524 CAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGC 3583

Db     72 GGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGACGCAGTTCCTGC 131

Qy   3584 TCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGA 3643

Db    132 ACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACA 191

Qy   3644 AGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCC 3703
```

```
               |   |          |||    ||| || | |||| |   ||||| |  | |   | | | || |
Db     192 ATCAAAGATGGAAGACACTCTAGAGCATACAGATTAAGAGGTGTCAGTGGAAACTGTATC 251

Qy    3704 CGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA 3739
           | ||||| |||||||||||| |   |||||||| |||
Db     252 CATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGA 287
```

Search completed: December 12, 2002, 16:39:17
Job time : 1217 secs

OM nucleic - nucleic search, using sw model

Run on: December 12, 2002, 14:13:30 ; Search time 166 Seconds
(without alignments)
11380.294 Million cell updates/sec

Title: US-09-902-432-3
Perfect score: 6160
Sequence: 1 cgggccggcgggagtagaag..........aaataaacatatttgactac 6160

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	4831.4	78.4	5134	2	US-08-635-121-1	Sequence 1, Appli
	2	2118.8	34.4	6605	1	US-08-769-309A-4	Sequence 4, Appli
	3	2118.8	34.4	6605	3	US-08-994-570-4	Sequence 4, Appli
c	4	99.8	1.6	7218	1	US-08-232-463-14	Sequence 14, Appl
c	5	71.2	1.2	1926	4	US-09-249-585A-4	Sequence 4, Appli
c	6	71.2	1.2	1931	2	US-09-130-114-2	Sequence 2, Appli
	7	70.8	1.1	1926	4	US-09-249-585A-2	Sequence 2, Appli
	8	70.8	1.1	2580	3	US-09-050-863-2	Sequence 2, Appli
	9	70.8	1.1	2580	4	US-09-359-081-2	Sequence 2, Appli
c	10	70.8	1.1	5452	2	US-09-130-114-1	Sequence 1, Appli
	11	70.8	1.1	9600	4	US-08-910-647-1	Sequence 1, Appli
	12	70.8	1.1	9600	4	US-09-620-925-1	Sequence 1, Appli
	13	70.8	1.1	10596	1	US-07-884-811-15	Sequence 15, Appl
	14	70.8	1.1	10596	1	US-07-885-971-15	Sequence 15, Appl
	15	70.8	1.1	10596	1	US-08-087-783A-15	Sequence 15, Appl
	16	70.8	1.1	10596	1	US-08-194-088B-15	Sequence 15, Appl
	17	70.8	1.1	10596	2	US-08-194-087-15	Sequence 15, Appl
	18	70.8	1.1	10596	5	PCT-US93-04648-15	Sequence 15, Appl
	19	66.4	1.1	3489	2	US-08-728-323A-1	Sequence 1, Appli
	20	66.4	1.1	3489	4	US-09-298-568-1	Sequence 1, Appli
c	21	66.4	1.1	32207	2	US-08-770-379-20	Sequence 20, Appl
c	22	66.4	1.1	32207	4	US-08-757-669A-20	Sequence 20, Appl
c	23	66.4	1.1	32207	4	US-09-230-371A-20	Sequence 20, Appl
c	24	64.6	1.0	16442	3	US-08-781-891-208	Sequence 208, App
	25	63.4	1.0	7218	1	US-08-232-463-14	Sequence 14, Appl
	26	60.4	1.0	2277	1	US-08-676-967-2	Sequence 2, Appli
	27	60.4	1.0	2277	1	US-08-676-974-2	Sequence 2, Appli
	28	60.4	1.0	2277	2	US-09-098-487-2	Sequence 2, Appli
	29	60.2	1.0	3337	1	US-08-072-610-1	Sequence 1, Appli
	30	60.2	1.0	3337	2	US-08-719-822B-1	Sequence 1, Appli
	31	60.2	1.0	3337	4	US-09-092-458-1	Sequence 1, Appli
c	32	58.8	1.0	51259	3	US-08-781-891-209	Sequence 209, App
	33	56.4	0.9	3211	2	US-08-574-959A-8	Sequence 8, Appli
	34	56.4	0.9	3211	4	US-09-357-014-8	Sequence 8, Appli
	35	56.4	0.9	3901	2	US-08-574-959A-6	Sequence 6, Appli
	36	56.4	0.9	3901	4	US-09-357-014-6	Sequence 6, Appli
	37	55	0.9	1052	1	US-08-466-603-1	Sequence 1, Appli
	38	55	0.9	1052	1	US-08-314-503A-1	Sequence 1, Appli
	39	55	0.9	1052	1	US-08-468-066-1	Sequence 1, Appli
	40	55	0.9	1052	2	US-08-466-717-1	Sequence 1, Appli
	41	55	0.9	1052	3	US-08-466-743-1	Sequence 1, Appli
	42	55	0.9	1052	5	PCT-US95-12414-1	Sequence 1, Appli
	43	52.6	0.9	1505	1	US-07-915-246-1	Sequence 1, Appli
	44	51.8	0.8	43795	3	US-08-742-185-101	Sequence 101, App
	45	47.8	0.8	2518	4	US-09-433-699-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-635-121-1
; Sequence 1, Application US/08635121
; Patent No. 5910442
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,121
;      FILING DATE:  19-APRIL-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/33603
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5134 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-635-121-1

  Query Match             78.4%;  Score 4831.4;  DB 2;  Length 5134;
  Best Local Similarity   97.9%;  Pred. No. 0;
```

```
Matches 5048;  Conservative    0;  Mismatches   76;  Indels   31;  Gaps   14;

Qy   1016 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 1075
Db      1 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 60

Qy   1076 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 1135
Db     61 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 120

Qy   1136 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 1195
Db    121 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 180

Qy   1196 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 1255
Db    181 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 240

Qy   1256 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 1315
Db    241 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 300

Qy   1316 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 1375
Db    301 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 360

Qy   1376 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 1435
Db    361 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 420

Qy   1436 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 1495
Db    421 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 480

Qy   1496 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 1555
Db    481 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 540

Qy   1556 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 1615
Db    541 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 600

Qy   1616 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 1675
Db    601 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 660

Qy   1676 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 1735
Db    661 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 720

Qy   1736 AGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGA 1795
Db    721 AGGGCCGCTGGAAGCA-CCCAGGATGGGGAAGCTGAGGAAGGAACTACTTC----GTGGA 775

Qy   1796 GAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 1855
Db    776 GAGAAGAAGAGGAAGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 835

Qy   1856 ACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 1915
Db    836 AC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 894

Qy   1916 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 1975
Db    895 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 954

Qy   1976 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 2035
Db    955 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 1014

Qy   2036 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGA 2095
Db   1015 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCA---GATAT 1071

Qy   2096 AGGAGGGCCAAGGACACTGGGA-GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 2154
Db   1072 AAGAGGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 1131

Qy   2155 AAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 2214
Db   1132 AAGAAGCC-GAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 1190

Qy   2215 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 2274
Db   1191 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 1250

Qy   2275 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCG 2334
Db   1251 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGAAG 1310

Qy   2335 AAGACTCTAGTGTAG--------AGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGA 2386
Db   1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy   2387 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA 2446
Db   1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy   2447 -GCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTA 2505
Db   1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy   2506 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAG 2565
Db   1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy   2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
Db   1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGT---AGTGTCCGAGGAGCTCA 1607

Qy   2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
Db   1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy   2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
Db   1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy   2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
Db   1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy   2806 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 2865
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy   2866 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 2925
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy   2926 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 2985
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy   2986 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 3045
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   3046 AGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 3105
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   3106 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 3165
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147
```

```
Qy   3166 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 3225
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   3226 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 3285
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   3286 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGG 3345
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   3346 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 3405
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   3406 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 3465
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   3466 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 3525
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   3526 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 3585
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   3586 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 3645
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   3646 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCG 3705
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   3706 TTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 3765
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   3766 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 3825
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   3826 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 3885
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   3886 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 3945
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   3946 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 4005
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   4006 AAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 4065
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   4066 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 4125
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   4126 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 4185
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   4186 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 4245
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   4246 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 4305
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   4306 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 4365
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   4366 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 4425
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   4426 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 4485
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   4486 CAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 4545
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   4546 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 4605
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   4606 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGG 4665
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   4666 AGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCT 4785
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGCTCAG--CTCGGAAATGCT 3765

Qy   4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 3825

Qy   4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
Db   3826 ACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 3884

Qy   4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
Db   3885 CAGAGGCA---GTGTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 3941

Qy   4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
Db   3942 TAACCGAGGAGCG-ATGCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 4000

Qy   5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 5085
Db   4001 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 4060

Qy   5086 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 5145
Db   4061 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 4120

Qy   5146 TCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATC 5205
Db   4121 TCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCATGAGATC 4179

Qy   5206 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 5265
Db   4180 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 4239

Qy   5266 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5325
Db   4240 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4299

Qy   5326 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5385
```

```
Db     4300 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4359

Qy     5386 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 5445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4360 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 4419

Qy     5446 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 5505
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4420 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 4479

Qy     5506 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 5565
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4480 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 4539

Qy     5566 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 5625
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4540 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 4599

Qy     5626 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 5685
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4600 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 4659

Qy     5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 5745
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4660 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 4719

Qy     5746 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATAT 5805
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4720 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATAT 4779

Qy     5806 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 5865
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4780 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 4839

Qy     5866 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 5925
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4840 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 4899

Qy     5926 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 5985
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4900 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 4959

Qy     5986 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 6045
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4960 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 5019

Qy     6046 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 6105
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5020 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 5079

Qy     6106 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5080 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 5134

RESULT 2
US-08-769-309A-4
; Sequence 4, Application US/08769309A
; Patent No. 5741890
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,309A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5741890and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6605 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  192..5531
US-08-769-309A-4

  Query Match             34.4%;  Score 2118.8;  DB 1;  Length 6605;
  Best Local Similarity   64.2%;  Pred. No. 0;
  Matches 4206;  Conservative    0;  Mismatches 1802;  Indels  546;  Gaps   46;

Qy        3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
            |||  |  |  | ||||| ||    ||||||||||||| || || ||||||||||||||
Db      162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy       63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
            ||||| |||||||||   | |||||||  |||||||   ||||  | || ||| |||
Db      222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy      120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
            || ||| | || || ||  ||  |         || |     | |  ||| | | |||||
Db      282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy      171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
            |||||||||||||||| |||||||||||| ||||||||  |  |||| |||||||||||
Db      342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy      231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
            ||||  ||  ||    |||| ||||   |||   | ||
Db      402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy      269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
                                | |||| ||||||||| |||    ||| ||||||||| ||
Db      462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy      309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
            || || ||||||||||| |||| ||||   | | |||| |||||  | || ||  | | |
Db      522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy      369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
            ||| | || |||||| | ||||||||||||||||       ||  || |||||||| |||
Db      582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy      429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
             ||||||||| |||| | |||||  |   ||| ||  |||||||||||||||||||| ||
Db      639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy      489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
            || || ||||| || ||||| |||||||| ||||||||||| ||||| |||||||| |
Db      699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy      549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
```

```
Db     759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy     609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db     819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy     669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db     864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy     729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db     924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy     789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db     981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy     849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db    1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy     909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db    1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy     969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db    1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy    1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db    1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy    1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db    1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy    1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db    1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy    1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db    1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy    1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db    1461 GAGGAGAGAACCGAAGAGCAGAAA------------------------ACGGAGGTGGAA 1496

Qy    1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db    1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy    1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db    1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy    1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db    1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy    1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db    1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy    1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db    1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy    1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db    1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy    1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db    1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy    1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db    1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy    1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db    1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy    1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db    2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy    1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db    2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy    1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db    2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy    2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db    2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy    2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db    2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy    2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db    2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy    2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db    2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy    2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db    2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy    2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db    2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy    2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db    2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy    2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db    2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy    2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db    2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy    2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db    2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy    2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db    2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy    2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db    2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933
```

```
Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGA---GGTGCATCTGGTACCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCT 3830

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3831 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3890

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3891 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3950

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3951 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4010

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4011 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4070

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4071 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4130

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db   4131 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4190

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4191 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4250

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4251 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4310

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4311 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4370

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4371 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4430

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4431 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4490

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4491 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4550

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4551 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4610

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4611 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4670

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4671 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4730

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4731 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4781

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4782 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4841

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4842 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4901

Qy   4617 GCTTATGATTCACAGACCCAGGTTC---------------CTGCATGCAGGCTTGACAG 4660
Db   4902 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4961

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4962 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5021

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5022 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5081

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
```

```
Db   5082 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5141

Qy   4819 ------------------------------------------------------------ 4818

Db   5142 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5201

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840

Db   5202 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5261

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898

Db   5262 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5321

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958

Db   5322 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5381

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018

Db   5382 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5441

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078

Db   5442 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5501

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117

Db   5502 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5561

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173

Db   5562 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5620

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233

Db   5621 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5679

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293

Db   5680 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5738

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353

Db   5739 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5778

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413

Db   5779 ------------------------------------------CATATAAATTCTGATTTA 5796

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472

Db   5797 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5856

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532

Db   5857 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5913

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575

Db   5914 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5973

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635

Db   5974 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6033

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686

Db   6034 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6093

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746

Db   6094 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6153

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806

Db   6154 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6213

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866

Db   6214 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6266

Qy   5867 TGTCAGA----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909

Db   6267 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6326

Qy   5910 GTGCCTAAGGCCATGAGA---------------TGTTTCCTAATGCAGAAGGAATCTGT 5953

Db   6327 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6386

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013

Db   6387 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6444

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073

Db   6445 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6504

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133

Db   6505 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6564

Qy   6134 CTGATTTAAATAAA 6147

Db   6565 CTGATTTAAATAAA 6578
```

RESULT 3
US-08-994-570-4
; Sequence 4, Application US/08994570
; Patent No. 6090929
; GENERAL INFORMATION:
;   APPLICANT: Scott, John D.,
;   APPLICANT: Nauert, Brian J.,
;   APPLICANT: Klauck, Theresa M.
;   TITLE OF INVENTION: Protein Binding Domains of Gravin
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower/233 South Wacker Drive
;   CITY: Chicago
;   STATE: Illinois
;   COUNTRY: United States of America
;   ZIP: 60606-6402
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/994,570
;   FILING DATE:
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: No. 6090929and, Greta E.
;   REGISTRATION NUMBER: 35,302
;   REFERENCE/DOCKET NUMBER: 27866/33451
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 312-474-6300
;   TELEFAX: 312-474-0448
;   TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 6605 base pairs
;   TYPE: nucleic acid

```
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  192..5531
US-08-994-570-4

  Query Match              34.4%;  Score 2118.8;  DB 3;  Length 6605;
  Best Local Similarity   64.2%;  Pred. No. 0;
  Matches 4206;  Conservative    0;  Mismatches 1802;  Indels  546;  Gaps   46;

Qy          3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db        162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy         63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db        222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy        120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db        282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy        171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db        342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy        231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG--------------------- 268
Db        402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy        269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db        462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy        309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db        522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy        369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db        582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy        429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db        639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy        489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db        699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy        549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db        759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy        609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db        819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy        669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db        864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy        729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db        924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy        789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db        981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy        849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db       1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy        909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db       1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy        969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db       1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy       1029 CCAGCCTCGGAGGA-------------------------GCAGGAGCCGGCAGAAGAC 1061
Db       1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy       1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db       1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy       1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db       1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy       1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db       1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy       1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db       1461 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1496

Qy       1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db       1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy       1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db       1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy       1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db       1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy       1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db       1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy       1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db       1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy       1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db       1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy       1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db       1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy       1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db       1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy       1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db       1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy       1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db       2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy       1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
```

```
Db    2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy    1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018

Db    2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy    2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078

Db    2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy    2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138

Db    2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy    2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198

Db    2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy    2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258

Db    2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy    2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318

Db    2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy    2319 CTGGAAGAGAAAGCCGAAGAC------------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366

Db    2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy    2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426

Db    2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy    2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486

Db    2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy    2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546

Db    2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy    2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597

Db    2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy    2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657

Db    2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy    2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717

Db    2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy    2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777

Db    2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy    2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834

Db    2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy    2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894

Db    3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy    2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954

Db    3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy    2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014

Db    3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy    3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074

Db    3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy    3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134

Db    3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy    3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191

Db    3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy    3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251

Db    3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy    3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311

Db    3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy    3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359

Db    3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy    3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401

Db    3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy    3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461

Db    3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy    3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521

Db    3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy    3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581

Db    3774 GAGGAGAATGA---GGTGCATCTGGTACCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCT 3830

Qy    3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641

Db    3831 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3890

Qy    3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701

Db    3891 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3950

Qy    3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752

Db    3951 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4010

Qy    3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806

Db    4011 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4070

Qy    3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866

Db    4071 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4130

Qy    3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911

Db    4131 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4190

Qy    3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971

Db    4191 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4250

Qy    3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028

Db    4251 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4310
```

```
Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4311 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4370

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4371 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4430

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4431 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4490

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4491 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4550

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4551 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4610

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4611 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4670

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4671 TTAAGTTCCGACCTGGAAGGAGAGAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4730

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4731 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4781

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4782 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4841

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4842 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4901

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4902 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4961

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4962 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5021

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5022 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5081

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5082 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5141

Qy   4819 ------------------------------------------------------------ 4818
Db   5142 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5201

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5202 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5261

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5262 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5321

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5322 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5381

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5382 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5441

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5442 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5501

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5502 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5561

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5562 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5620

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5621 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5679

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5680 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5738

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5739 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5778

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5779 -----------------------------------------CATATAAATTCTGATTTA 5796

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5797 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5856

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5857 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5913

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5914 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5973

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5974 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6033

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6034 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6093

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6094 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6153

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6154 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTGAAACATACAGAATGTTCTATTGTC 6213

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6214 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6266

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6267 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6326

Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6327 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6386

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
```

```
Db    6387 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6444

Qy    6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db    6445 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6504

Qy    6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db    6505 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6564

Qy    6134 CTGATTTAAATAAA 6147
Db    6565 CTGATTTAAATAAA 6578
```

RESULT 4
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
US-08-232-463-14

```
  Query Match             1.6%;  Score 99.8;  DB 1;  Length 7218;
  Best Local Similarity   2.1%;  Pred. No. 1.5e-15;
  Matches    8;  Conservative  262;  Mismatches  109;  Indels    0;  Gaps    0;

Qy     944 AAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGA 1003
Db    1431 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1372

Qy    1004 CGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACAC 1063
Db    1371 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1312

Qy    1064 AGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGT 1123
Db    1311 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1252

Qy    1124 TGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGA 1183
Db    1251 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1192

Qy    1184 TGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAA 1243
Db    1191 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1132

Qy    1244 GACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAAC 1303
Db    1131 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1072

Qy    1304 AGGAGAATCCTTGCCCCCT 1322
Db    1071 RRRRRATCGCAAGCTCCCT 1053
```

RESULT 5
US-09-249-585A-4/c
; Sequence 4, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 1926
; TYPE: DNA
; ORGANISM: Epstein Barr Virus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(1926)
; OTHER INFORMATION: template strand of EBNA-1 DNA
US-09-249-585A-4

```
  Query Match             1.2%;  Score 71.2;  DB 4;  Length 1926;
  Best Local Similarity   47.7%;  Pred. No. 1.9e-08;
  Matches  208;  Conservative    0;  Mismatches  228;  Indels    0;  Gaps    0;

Qy     980 GGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGA 1039
Db     977 GGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGA 918

Qy    1040 GGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGT 1099
Db     917 GGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGG 858

Qy    1100 GGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGC 1159
Db     857 GGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGA 798

Qy    1160 TCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGA 1219
Db     797 CGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGA 738

Qy    1220 GGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGA 1279
Db     737 GGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGG 678

Qy    1280 AGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCC 1339
Db     677 GGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGG 618
```

```
Qy    1340 CCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTG 1399
Db     617 GGAGGACGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGA 558

Qy    1400 TGTCTCTGGAGGAGAC 1415
Db     557 CGAGGACGGGGAGGAC 542
```

RESULT 6
US-09-130-114-2/c
; Sequence 2, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert A.
; APPLICANT: Damaj, Bassam B.
; APPLICANT: Robbins, Alan K.
; TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
; TITLE OF INVENTION: From Multiple Transfected Episomes
; FILE REFERENCE: 0867/1D903US1
; CURRENT APPLICATION NUMBER: US/09/130,114
; CURRENT FILING DATE: 1998-08-06
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 1931
;   TYPE: DNA
;   ORGANISM: EBNA
US-09-130-114-2

```
  Query Match             1.2%;  Score 71.2;  DB 2;  Length 1931;
  Best Local Similarity   47.7%;  Pred. No. 1.9e-08;
  Matches  208;  Conservative    0;  Mismatches  228;  Indels    0;  Gaps    0;

Qy     980 GGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGA 1039
Db     977 GGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGA 918

Qy    1040 GGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGT 1099
Db     917 GGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGG 858

Qy    1100 GGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGC 1159
Db     857 GGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGAGGACGAGGACGGGGAGGA 798

Qy    1160 TCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGA 1219
Db     797 CGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGA 738

Qy    1220 GGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGA 1279
Db     737 GGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGG 678

Qy    1280 AGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCC 1339
Db     677 GGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGAGGACGGGGAGGACGGGGAGGACGG 618

Qy    1340 CCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTG 1399
Db     617 GGAGGACGAGGACGGGGAGGAGGACGAGGACGGGGAGGACGGGGAGGACGGGGAGGAGGA 558

Qy    1400 TGTCTCTGGAGGAGAC 1415
Db     557 CGAGGACGGGGAGGAC 542
```

RESULT 7
US-09-249-585A-2
; Sequence 2, Application US/09249585A
; Patent No. 6417002
; GENERAL INFORMATION:
; APPLICANT: Horlick, Robert
; TITLE OF INVENTION: METHOD FOR MAINTENANCE AND SELECTION OF EPISOMES
; FILE REFERENCE: 0867/0D905
; CURRENT APPLICATION NUMBER: US/09/249,585A
; CURRENT FILING DATE: 1999-02-11
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
;   LENGTH: 1926
;   TYPE: DNA
;   ORGANISM: Epstein Barr Virus
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(1926)
;   OTHER INFORMATION: coding strand of EBNA-1 DNA
US-09-249-585A-2

```
  Query Match             1.1%;  Score 70.8;  DB 4;  Length 1926;
  Best Local Similarity   45.3%;  Pred. No. 2.4e-08;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy     946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db     333 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 392

Qy    1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db     393 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 452

Qy    1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db     453 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 512

Qy    1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db     513 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 572

Qy    1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db     573 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 632

Qy    1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db     633 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 691

Qy    1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db     692 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 749

Qy    1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db     750 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 809

Qy    1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db     810 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 869

Qy    1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db     870 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 929

Qy    1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
Db     930 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 989

Qy    1606 GTGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665
Db     990 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 1049

Qy    1666 CTGATGAGCAGAAGGGAGAG 1685
Db    1050 TGGAGGCCGCCGGGGTAGAG 1069
```

```
RESULT 8
US-09-050-863-2
; Sequence 2, Application US/09050863
; Patent No. 6114111
;  GENERAL INFORMATION:
;    APPLICANT:  Lao, Ying
;    APPLICANT:  Hiang, Betty
;    APPLICANT:  Payan, Don
;    TITLE OF INVENTION:  Mammalian Protein Interaction Cloning
;    TITLE OF INVENTION:  System
;    NUMBER OF SEQUENCES:  5
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Flehr, Hohbach, Test, Albritton & Herbert
;      STREET:  4 Embarcadero Center, Suite 3400
;      CITY:  San Francisco
;      STATE:  CA
;      COUNTRY:  USA
;      ZIP:  94111-4187
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/050,863
;      FILING DATE:  30-MAR-1998
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Silva, Robin M.
;      REGISTRATION NUMBER:  38,304
;      REFERENCE/DOCKET NUMBER:  A-65638/DJB/RMS
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (415) 781-1989
;      TELEFAX:  (415) 949-8711
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2580 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  unknown
;      TOPOLOGY:  unknown
;    MOLECULE TYPE:  DNA
US-09-050-863-2

  Query Match             1.1%;  Score 70.8;  DB 3;  Length 2580;
  Best Local Similarity   45.3%;  Pred. No. 2.8e-08;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy    946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db    716 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 775

Qy   1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db    776 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 835

Qy   1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db    836 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 895

Qy   1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db    896 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 955

Qy   1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db    956 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1015

Qy   1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db   1016 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1074

Qy   1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db   1075 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 1132

Qy   1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db   1133 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1192

Qy   1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db   1193 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 1252

Qy   1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db   1253 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 1312

Qy   1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
Db   1313 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 1372

Qy   1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665
Db   1373 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 1432

Qy   1666 CTGATGAGCAGAAGGGAGAG 1685
Db   1433 TGGAGGCCGCCGGGGTAGAG 1452


RESULT 9
US-09-359-081-2
; Sequence 2, Application US/09359081
; Patent No. 6316223
;  GENERAL INFORMATION:
;       APPLICANT: Lao, Ying
;                  Hiang, Betty
;                  Payan, Don
;       TITLE OF INVENTION: Mammalian Protein Interaction Cloning
;                           System
;       NUMBER OF SEQUENCES: 5
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
;            STREET: 4 Embarcadero Center, Suite 3400
;            CITY: San Francisco
;            STATE: CA
;            COUNTRY: USA
;            ZIP: 94111-4187
;       COMPUTER READABLE FORM:
;            MEDIUM TYPE: Floppy disk
;            COMPUTER: IBM PC compatible
;            OPERATING SYSTEM: PC-DOS/MS-DOS
;            SOFTWARE: PatentIn Release #1.0, Version #1.30
;       CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/359,081
;            FILING DATE: 22-Jul-1999
;            CLASSIFICATION: <Unknown>
;       PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 09/050,863
;            FILING DATE: <Unknown>
;       ATTORNEY/AGENT INFORMATION:
;            NAME: Silva, Robin M.
;            REGISTRATION NUMBER: 38,304
;            REFERENCE/DOCKET NUMBER: A-65638/DJB/RMS
;       TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (415) 781-1989
;            TELEFAX: (415) 949-8711
;  INFORMATION FOR SEQ ID NO: 2:
;       SEQUENCE CHARACTERISTICS:
;            LENGTH: 2580 base pairs
;            TYPE: nucleic acid
;            STRANDEDNESS: unknown
;            TOPOLOGY: unknown
;       MOLECULE TYPE: DNA
;       SEQUENCE DESCRIPTION: SEQ ID NO: 2:
```

```
US-09-359-081-2

  Query Match             1.1%;  Score 70.8;  DB 4;  Length 2580;
  Best Local Similarity   45.3%;  Pred. No. 2.8e-08;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy     946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db     716 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 775

Qy    1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db     776 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 835

Qy    1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db     836 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 895

Qy    1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db     896 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 955

Qy    1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db     956 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1015

Qy    1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db    1016 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1074

Qy    1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db    1075 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 1132

Qy    1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db    1133 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1192

Qy    1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db    1193 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 1252

Qy    1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db    1253 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 1312

Qy    1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
Db    1313 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 1372

Qy    1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665
Db    1373 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 1432

Qy    1666 CTGATGAGCAGAAGGGAGAG 1685
Db    1433 TGGAGGCCGCCGGGGTAGAG 1452


RESULT 10
US-09-130-114-1/c
; Sequence 1, Application US/09130114
; Patent No. 5976807
; GENERAL INFORMATION:
;  APPLICANT: Horlick, Robert A.
;  APPLICANT: Damaj, Bassam B.
;  APPLICANT: Robbins, Alan K.
;  TITLE OF INVENTION: Eukaryotic Cells Stably Expressing Genes
;  TITLE OF INVENTION: From Multiple Transfected Episomes
;  FILE REFERENCE: 0867/1D903US1
;  CURRENT APPLICATION NUMBER: US/09/130,114
;  CURRENT FILING DATE: 1998-08-06
;  NUMBER OF SEQ ID NOS: 36
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1
;   LENGTH: 5452
;   TYPE: DNA
;   ORGANISM: VEBNA
US-09-130-114-1

  Query Match             1.1%;  Score 70.8;  DB 2;  Length 5452;
  Best Local Similarity   45.3%;  Pred. No. 4.3e-08;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy     946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db    2089 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 2030

Qy    1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db    2029 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 1970

Qy    1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db    1969 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 1910

Qy    1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db    1909 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 1850

Qy    1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db    1849 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1790

Qy    1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db    1789 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1731

Qy    1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db    1730 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 1673

Qy    1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db    1672 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1613

Qy    1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db    1612 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 1553

Qy    1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db    1552 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 1493

Qy    1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
Db    1492 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 1433

Qy    1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665
Db    1432 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 1373

Qy    1666 CTGATGAGCAGAAGGGAGAG 1685
Db    1372 TGGAGGCCGCCGGGGTAGAG 1353


RESULT 11
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
;  GENERAL INFORMATION:
;    APPLICANT:  Zuckermann et al.
;    TITLE OF INVENTION:  Compositions and Methods for
;    TITLE OF INVENTION:  Polynucleotide Delivery
;    NUMBER OF SEQUENCES:  4
```

```
;     CORRESPONDENCE ADDRESS:
;       ADDRESSEE:  Chiron Corporation
;       STREET:  4560 Horton Street
;       CITY:  Emeryville
;       STATE:  California
;       COUNTRY:  U.S.A.
;       ZIP:  94608-2916
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/910,647
;       FILING DATE:
;       CLASSIFICATION:  514
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  Fujita, Sharon M.
;       REGISTRATION NUMBER:  38,459
;       REFERENCE/DOCKET NUMBER:  1218.002
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  (510) 923-2706
;       TELEFAX:  (510) 655-3542
;   INFORMATION FOR SEQ ID NO:  1:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  9600 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  DNA (genomic)
US-08-910-647-1

  Query Match             1.1%;  Score 70.8;  DB 4;  Length 9600;
  Best Local Similarity   45.3%;  Pred. No. 6e-08;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy     946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005

Db     762 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 821

Qy    1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065

Db     822 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 881

Qy    1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125

Db     882 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 941

Qy    1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185

Db     942 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 1001

Qy    1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245

Db    1002 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1061

Qy    1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305

Db    1062 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1120

Qy    1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365

Db    1121 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 1178

Qy    1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425

Db    1179 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1238

Qy    1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485

Db    1239 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 1298

Qy    1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545

Db    1299 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 1358

Qy    1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605

Db    1359 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 1418

Qy    1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665

Db    1419 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 1478

Qy    1666 CTGATGAGCAGAAGGGAGAG 1685

Db    1479 TGGAGGCCGCCGGGGTAGAG 1498


RESULT 12
US-09-620-925-1
; Sequence 1, Application US/09620925
; Patent No. 6468986
;   GENERAL INFORMATION:
;         APPLICANT: Zuckermann et al.
;         TITLE OF INVENTION: Compositions and Methods for
;                             Polynucleotide Delivery
;         NUMBER OF SEQUENCES: 4
;         CORRESPONDENCE ADDRESS:
;              ADDRESSEE: Chiron Corporation
;              STREET: 4560 Horton Street
;              CITY: Emeryville
;              STATE: California
;              COUNTRY: U.S.A.
;              ZIP: 94608-2916
;         COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;         CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/09/620,925
;              FILING DATE: 21-Jul-2000
;              CLASSIFICATION: &lt;Unknown&gt;
;         PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/910,647
;              FILING DATE: &lt;Unknown&gt;
;         ATTORNEY/AGENT INFORMATION:
;              NAME: Fujita, Sharon M.
;              REGISTRATION NUMBER: 38,459
;              REFERENCE/DOCKET NUMBER: 1218.002
;         TELECOMMUNICATION INFORMATION:
;              TELEPHONE: (510) 923-2706
;              TELEFAX: (510) 655-3542
;   INFORMATION FOR SEQ ID NO: 1:
;         SEQUENCE CHARACTERISTICS:
;              LENGTH: 9600 base pairs
;              TYPE: nucleic acid
;              STRANDEDNESS: single
;              TOPOLOGY: linear
;         MOLECULE TYPE: DNA (genomic)
;         SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-620-925-1

  Query Match             1.1%;  Score 70.8;  DB 4;  Length 9600;
  Best Local Similarity   45.3%;  Pred. No. 6e-08;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy     946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005

Db     762 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 821

Qy    1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065

Db     822 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 881
```

```
Qy   1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db    882 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 941

Qy   1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db    942 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 1001

Qy   1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db   1002 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1061

Qy   1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db   1062 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 1120

Qy   1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db   1121 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 1178

Qy   1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db   1179 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 1238

Qy   1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db   1239 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 1298

Qy   1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db   1299 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 1358

Qy   1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
Db   1359 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 1418

Qy   1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665
Db   1419 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 1478

Qy   1666 CTGATGAGCAGAAGGGAGAG 1685
Db   1479 TGGAGGCCGCCGGGGTAGAG 1498
```

RESULT 13
US-07-884-811-15
; Sequence 15, Application US/07884811
; Patent No. 5316921
; GENERAL INFORMATION:
; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/884,811
; FILING DATE: 19920518
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 755.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10596 bases
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
US-07-884-811-15

Query Match 1.1%; Score 70.8; DB 1; Length 10596;
Best Local Similarity 45.3%; Pred. No. 6.3e-08;
Matches 335; Conservative 0; Mismatches 402; Indels 3; Gaps 2;

```
Qy    946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db   2296 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 2355

Qy   1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db   2356 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 2415

Qy   1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db   2416 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 2475

Qy   1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db   2476 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 2535

Qy   1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db   2536 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 2595

Qy   1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db   2596 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2654

Qy   1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db   2655 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 2712

Qy   1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db   2713 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 2772

Qy   1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db   2773 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 2832

Qy   1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db   2833 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 2892

Qy   1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
Db   2893 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 2952

Qy   1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665
Db   2953 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 3012

Qy   1666 CTGATGAGCAGAAGGGAGAG 1685
Db   3013 TGGAGGCCGCCGGGGTAGAG 3032
```

RESULT 14

```
US-07-885-971-15
; Sequence 15, Application US/07885971
; Patent No. 5328837
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J.  Lokker, Nathalie A.  Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS
;    NUMBER OF SEQUENCES:  21
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  5.25 inch, 360 Kb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  patin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/885,971
;      FILING DATE:  19920518
;      CLASSIFICATION:  530
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Dreger, Ginger R.
;      REGISTRATION NUMBER:  33,055
;      REFERENCE/DOCKET NUMBER:  779
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-3216
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 bases
;      TYPE:  NUCLEIC ACID
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-07-885-971-15

  Query Match             1.1%;  Score 70.8;  DB 1;  Length 10596;
  Best Local Similarity   45.3%;  Pred. No. 6.3e-08;
  Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy    946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005

Db   2296 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 2355

Qy   1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065

Db   2356 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 2415

Qy   1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125

Db   2416 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 2475

Qy   1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185

Db   2476 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 2535

Qy   1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245

Db   2536 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 2595

Qy   1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305

Db   2596 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2654

Qy   1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365

Db   2655 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 2712

Qy   1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425

Db   2713 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 2772

Qy   1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485

Db   2773 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 2832

Qy   1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545

Db   2833 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 2892

Qy   1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605

Db   2893 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 2952

Qy   1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665

Db   2953 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 3012

Qy   1666 CTGATGAGCAGAAGGGAGAG 1685

Db   3013 TGGAGGCCGCCGGGGTAGAG 3032


RESULT 15
US-08-087-783A-15
; Sequence 15, Application US/08087783A
; Patent No. 5547856
;  GENERAL INFORMATION:
;    APPLICANT:  Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R.
;    TITLE OF INVENTION:  HEPATOCYTE GROWTH FACTOR VARIANTS
;    NUMBER OF SEQUENCES:  22
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Genentech, Inc.
;      STREET:  460 Point San Bruno Blvd
;      CITY:  South San Francisco
;      STATE:  California
;      COUNTRY:  USA
;      ZIP:  94080
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  3.5 inch, 1.44 Mb floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  WinPatin (Genentech)
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/087,783A
;      FILING DATE:  13-Jul-1993
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/884811
;      FILING DATE:  18-MAY-92
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  07/885971
;      FILING DATE:  18-MAY-92
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Marschang, Diane L.
;      REGISTRATION NUMBER:  35,600
;      REFERENCE/DOCKET NUMBER:  P0755779P1
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  415/225-5416
;      TELEFAX:  415/952-9881
;      TELEX:  910/371-7168
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  10596 base pairs
;      TYPE:  Nucleic Acid
;      STRANDEDNESS:  Single
;      TOPOLOGY:  Linear
US-08-087-783A-15

  Query Match             1.1%;  Score 70.8;  DB 1;  Length 10596;
```

```
Best Local Similarity   45.3%;  Pred. No. 6.3e-08;
Matches  335;  Conservative    0;  Mismatches  402;  Indels    3;  Gaps    2;

Qy     946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db    2296 AGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 2355

Qy    1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db    2356 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGG 2415

Qy    1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db    2416 AGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAGGAGC 2475

Qy    1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATG 1185
Db    2476 AGGAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGG 2535

Qy    1186 AGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGA 1245
Db    2536 AGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 2595

Qy    1246 CAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAG 1305
Db    2596 AGGGGCAGGAGCA-GGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGGGGCAG 2654

Qy    1306 GAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGC 1365
Db    2655 GAGCA--GGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGGGC 2712

Qy    1366 CTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAAC 1425
Db    2713 AGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGG 2772

Qy    1426 TGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTG 1485
Db    2773 AGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGAGGGGCAGGAGGGGCAGGAGCAGG 2832

Qy    1486 AGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAAC 1545
Db    2833 AGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGGGGCAGGAGCAGGAGGGGCAGGAGG 2892

Qy    1546 TCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAG 1605
Db    2893 GGCAGGAGCAGGAGGAGGGGCAGGAGCAGGAGGGGCAGGAGCAGGAGGTGGAGGCCGGGG 2952

Qy    1606 GTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTG 1665
Db    2953 TCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGAGGCCGGGGTCGAGGAGGTAG 3012

Qy    1666 CTGATGAGCAGAAGGGAGAG 1685
Db    3013 TGGAGGCCGCCGGGGTAGAG 3032

Search completed: December 12, 2002, 14:19:35
Job time : 351 secs
```

GenCore version 5.1.3

OM nucleic - nucleic search, using sw model

```
Run on:          December 12, 2002, 16:19:50 ; Search time 244 Seconds
                                  (without alignments)
                                  9983.770 Million cell updates/sec

Title:           US-09-902-432-3
Perfect score:   6160
Sequence:        1 cgggccggcgggagtagaag..........aaataaacatatttgactac 6160

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        355320 seqs, 197730502 residues

Total number of hits satisfying chosen parameters:      710640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_NA:*
                 1:  /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
                 2:  /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 3:  /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 4:  /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                 5:  /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 6:  /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                 7:  /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 8:  /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                 9:  /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                 10:  /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
                 11:  /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                 12:  /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
                 13:  /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                 14:  /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6160	100.0	6160	9	US-09-902-432-3	Sequence 3, Appli
	2	4831.4	78.4	5134	9	US-09-902-432-1	Sequence 1, Appli
	3	2144.4	34.8	6608	10	US-09-738-877-1	Sequence 1, Appli
	4	2144.4	34.8	6608	10	US-09-880-107-3439	Sequence 3439, Ap
	5	1945.6	31.6	5346	10	US-09-738-877-2	Sequence 2, Appli
c	6	550	8.9	563	10	US-09-917-800A-1223	Sequence 1223, Ap
	7	162.2	2.6	178	10	US-09-917-800A-1475	Sequence 1475, Ap
	8	106	1.7	720	9	US-10-101-487-74	Sequence 74, Appl
c	9	106	1.7	720	9	US-10-101-487-76	Sequence 76, Appl
	10	89.6	1.5	522	9	US-10-101-487-71	Sequence 71, Appl
c	11	89.6	1.5	530	9	US-10-101-487-73	Sequence 73, Appl
	12	89.6	1.5	554	9	US-10-101-487-69	Sequence 69, Appl
	13	89.6	1.5	554	9	US-10-101-487-106	Sequence 106, App
	14	71	1.2	575	10	US-09-864-761-20733	Sequence 20733, A
	15	71	1.2	1969	10	US-09-864-761-3972	Sequence 3972, Ap
c	16	65.8	1.1	659158	9	US-09-771-208-20	Sequence 20, Appl
	17	64.4	1.0	100	10	US-09-728-445-336	Sequence 336, App
	18	61	1.0	700	10	US-09-864-761-17529	Sequence 17529, A
c	19	59.4	1.0	53226	10	US-09-818-264-3	Sequence 3, Appli
c	20	59.2	1.0	511	10	US-09-864-761-22817	Sequence 22817, A
c	21	57.2	0.9	475	10	US-09-864-761-6203	Sequence 6203, Ap
	22	56.2	0.9	474	10	US-09-864-761-3470	Sequence 3470, Ap
	23	55.4	0.9	2142	10	US-09-917-800A-1567	Sequence 1567, Ap
	24	54.2	0.9	334	10	US-09-864-761-20240	Sequence 20240, A
c	25	53.8	0.9	267	10	US-09-864-761-27984	Sequence 27984, A
c	26	53.8	0.9	474	10	US-09-864-761-11284	Sequence 11284, A
	27	53.8	0.9	3809	12	US-10-001-870-68	Sequence 68, Appl
	28	52.8	0.9	726	10	US-09-864-761-19707	Sequence 19707, A
	29	52.8	0.9	1951	10	US-09-864-761-2926	Sequence 2926, Ap
c	30	52.4	0.9	5762	10	US-09-835-232-3	Sequence 3, Appli
	31	50.6	0.8	2623	9	US-09-804-409A-7	Sequence 7, Appli
c	32	50.4	0.8	5387	9	US-10-001-873-22	Sequence 22, Appl
c	33	50.2	0.8	438	10	US-09-864-761-4988	Sequence 4988, Ap
	34	49.6	0.8	663	9	US-10-001-857-11	Sequence 11, Appl
	35	49.4	0.8	659158	9	US-09-771-208-20	Sequence 20, Appl
	36	49	0.8	1728	9	US-09-906-514-3	Sequence 3, Appli
c	37	48.8	0.8	360	10	US-09-864-761-28274	Sequence 28274, A
c	38	48.8	0.8	450	10	US-09-864-761-11696	Sequence 11696, A
	39	48.4	0.8	700	9	US-10-040-739-1385	Sequence 1385, Ap
	40	48.4	0.8	768	9	US-09-938-842A-812	Sequence 812, App
	41	48.4	0.8	4316	10	US-09-880-107-3713	Sequence 3713, Ap
c	42	48	0.8	276	10	US-09-864-761-20595	Sequence 20595, A
	43	48	0.8	342	10	US-09-822-263-11	Sequence 11, Appl
	44	48	0.8	555	10	US-09-822-263-3	Sequence 3, Appli
c	45	48	0.8	742	10	US-09-770-149-52	Sequence 52, Appl

ALIGNMENTS

```
RESULT 1
US-09-902-432-3
; Sequence 3, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 6160
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-902-432-3

  Query Match            100.0%;  Score 6160;  DB 9;  Length 6160;
  Best Local Similarity  100.0%;  Pred. No. 0;
  Matches 6160;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGC 60

Qy         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 GGAGCCCCGAGCAGCCGGCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATG 120

Qy        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 GGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCA 180

Qy        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 AGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATG 240

Qy        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TCCATGTCCAAGAGGAAAACCAGGAGGGGCAGGAGGAAGAAGTCGTTGATGAGGATGTTG 300

Qy        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACT 360

Qy        361 CCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAAC 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATTGAAC 420

Qy        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 AGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTA 480

Qy        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 ATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGG 540

Qy        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 ATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAG 600

Qy        601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCG 660

Qy        661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 GAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCC 720

Qy        721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 TGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTG 780

Qy        781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 AGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAG 840

Qy        841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 AATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTC 900

Qy        901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 ACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGG 960

Qy        961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 AAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAA 1020

Qy       1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGT 1080

Qy       1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 CAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCAT 1140

Qy       1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 CGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCC 1200

Qy       1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 ACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGG 1260

Qy       1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCC 1320

Qy       1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGA 1380

Qy       1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 TGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTC 1440

Qy       1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 CTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGT 1500

Qy       1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAG 1560

Qy       1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 GCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAG 1620

Qy       1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1621 AGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGG 1680

Qy       1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1681 GAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGC 1740

Qy       1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 CGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGA 1800

Qy       1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 AGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGG 1860

Qy       1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 TCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCT 1920

Qy       1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 TGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGG 1980

Qy       1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 AACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCACTGA 2040

Qy       2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2041 TTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAG 2100

Qy       2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2101 GGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAG 2160

Qy       2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2161 CCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCT 2220

Qy       2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2221 CACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTA 2280

Qy       2281 AAAGATTAGTCACTCCAAGAAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2281 AAAGATTAGTCACTCCAAGAAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGACT 2340

Qy       2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2341 CTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCA 2400
```

```
Qy   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCA 2460

Qy   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 CTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCCGTCG 2520

Qy   2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 TGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAATACGG 2580

Qy   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 AGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCC 2640

Qy   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 ACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGT 2700

Qy   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCA 2760

Qy   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 TGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCA 2820

Qy   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 CCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATT 2880

Qy   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTA 2940

Qy   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 CCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTG 3000

Qy   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 ACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAG 3060

Qy   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGA 3120

Qy   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 AAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGG 3180

Qy   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 AGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACG 3240

Qy   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 TTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGA 3300

Qy   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGGCAGATGTAGACCATG 3360

Qy   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 TCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGT 3420

Qy   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 CATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACA 3480

Qy   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3481 CTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAG 3540

Qy   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 CTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTT 3600

Qy   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 CGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATG 3660

Qy   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 TTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTG 3720

Qy   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 AGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGG 3780

Qy   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGACAGTG 3840

Qy   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 AAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTTAGCC 3900

Qy   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 TGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGC 3960

Qy   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 AAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGA 4020

Qy   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 AGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTG 4080

Qy   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140

Qy   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200

Qy   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260

Qy   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320

Qy   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380

Qy   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440

Qy   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500

Qy   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560

Qy   4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620

Qy     4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680

Qy     4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4681 GGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGACTTGC 4740

Qy     4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4741 AAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGT 4800

Qy     4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4801 TGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCAACCCAAAGATCCAAA 4860

Qy     4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4861 AGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGGCAGAGGCCAGTGCCT 4920

Qy     4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4921 CTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGCTAACCGAGGAGGGCG 4980

Qy     4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     4981 ATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGTCAGTCAAAGAGAACA 5040

Qy     5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5041 AGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGGCAGAATCCTAAGATG 5100

Qy     5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5101 TTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAAGATG 5160

Qy     5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5161 CTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCG 5220

Qy     5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5221 TCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGA 5280

Qy     5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5281 GCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5340

Qy     5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5341 TATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTG 5400

Qy     5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5401 TATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCTGCCATACCTAGTTCC 5460

Qy     5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5461 ACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCC 5520

Qy     5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTGTATTATGCAA 5580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5521 TCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTGTATTATGCAA 5580

Qy     5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5581 TATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACC 5640

Qy     5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5641 TGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGAT 5700

Qy     5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5701 GCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCT 5760

Qy     5761 CAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTC 5820
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5761 CAGAAATAAAAACCCCATTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTC 5820

Qy     5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5821 CCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCC 5880

Qy     5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5881 CTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGC 5940

Qy     5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     5941 AGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATAT 6000

Qy     6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTGATAGATACTGGAGTTTGTC 6060
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6001 GCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTGATAGATACTGGAGTTTGTC 6060

Qy     6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     6061 TGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAAT 6120

Qy     6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
            ||||||||||||||||||||||||||||||||||||||||
Db     6121 TGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160

RESULT 2
US-09-902-432-1
; Sequence 1, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 5134
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-902-432-1

  Query Match             78.4%;  Score 4831.4;  DB 9;  Length 5134;
  Best Local Similarity   97.9%;  Pred. No. 0;
  Matches 5048;  Conservative    0;  Mismatches   76;  Indels   31;  Gaps   14;

Qy     1016 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 1075
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 GGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAG 60

Qy     1076 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 1135
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 GTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGA 120

Qy     1136 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 1195
```

```
Db    121 GGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGA 180

Qy   1196 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 1255
Db    181 AGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGA 240

Qy   1256 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 1315
Db    241 GCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTT 300

Qy   1316 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 1375
Db    301 GCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGA 360

Qy   1376 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 1435
Db    361 GCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCT 420

Qy   1436 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 1495
Db    421 AAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGAT 480

Qy   1496 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 1555
Db    481 GCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAG 540

Qy   1556 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 1615
Db    541 CTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGA 600

Qy   1616 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 1675
Db    601 CGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCA 660

Qy   1676 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 1735
Db    661 GAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAA 720

Qy   1736 AGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGA 1795
Db    721 AGGGCCGCTGGAAGCA-CCCAGGATGGGGAAGCTGAGGAAGGAACTACTTC----GTGGA 775

Qy   1796 GAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 1855
Db    776 GAGAAGAAGAGGAAGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAA 835

Qy   1856 ACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 1915
Db    836 AC-GGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGC 894

Qy   1916 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 1975
Db    895 CACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGG 954

Qy   1976 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 2035
Db    955 TGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGC 1014

Qy   2036 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGA 2095
Db   1015 ACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCA---GATAT 1071

Qy   2096 AGGAGGGCCAAGGACACTGGGA-GGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 2154
Db   1072 AAGAGGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACA 1131

Qy   2155 AAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 2214
Db   1132 AAGAAGCC-GAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAA 1190

Qy   2215 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 2274
Db   1191 GTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGT 1250

Qy   2275 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCG 2334
Db   1251 CATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGAAG 1310

Qy   2335 AAGACTCTAGTGTAG--------AGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGA 2386
Db   1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy   2387 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAA 2446
Db   1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy   2447 -GCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTA 2505
Db   1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy   2506 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGGAAG 2565
Db   1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy   2566 CCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCA 2625
Db   1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGT---AGTGTCCGAGGAGCTCA 1607

Qy   2626 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 2685
Db   1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy   2686 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 2745
Db   1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy   2746 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 2805
Db   1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy   2806 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 2865
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy   2866 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 2925
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy   2926 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 2985
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy   2986 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 3045
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   3046 AGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 3105
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   3106 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 3165
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   3166 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 3225
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   3226 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 3285
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   3286 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACGG 3345
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327
```

```
Qy   3346 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 3405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   3406 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 3465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   3466 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 3525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   3526 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 3585
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   3586 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 3645
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   3646 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCG 3705
          ||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   3706 TTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 3765
          |   |  |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   3766 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 3825
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   3826 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 3885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   3886 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 3945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   3946 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 4005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   4006 AAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 4065
          || | | |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   4066 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 4125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   4126 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 4185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   4186 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 4245
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   4246 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 4305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   4306 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 4365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   4366 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 4425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   4426 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 4485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   4486 CAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 4545
          ||||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   4546 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 4605
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   4606 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGG 4665
          ||||||||||||||||||||||||||||||||||||||||     |  | ||||||||||
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGACAGCAGGG 3647

Qy   4666 AGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
          |||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCT 4785
          ||||||||||||||||||||||||||||||||||||||||          ||||||||||
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGCTCAG--CTCGGAAATGCT 3765

Qy   4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 3825

Qy   4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
          |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   3826 ACCCAAAGAT-CAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 3884

Qy   4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
          ||||||     |  ||||||||||||||||||||||||||||||||||||||||||||||
Db   3885 CAGAGGCA---GTGTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 3941

Qy   4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
          |||||||||| |      ||||||||||||||||||||||||||||||||||||||||||
Db   3942 TAACCGAGGAGCG-ATGCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 4000

Qy   5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 5085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4001 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGAGACCTGG 4060

Qy   5086 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 5145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4061 CAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAG 4120

Qy   5146 TCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATC 5205
          |||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db   4121 TCACAGAACAAGATGCTGCTGTTGGGACCTTG-GACCAAGATTTCAGAGCCCATGAGATC 4179

Qy   5206 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 5265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4180 CAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCT 4239

Qy   5266 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4240 TCATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4299

Qy   5326 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 5385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4300 TTGATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCC 4359

Qy   5386 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 5445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4360 TTGATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATTCT 4419

Qy   5446 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 5505
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4420 GCCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGT 4479

Qy   5506 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCT 5565
```

```
Db    4480 TTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTCTAACGTTTAAGCACATGCT 4539

Qy    5566 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 5625
Db    4540 TTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAA 4599

Qy    5626 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 5685
Db    4600 GCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGACAAGTCTTTT 4659

Qy    5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 5745
Db    4660 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCC 4719

Qy    5746 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 5805
Db    4720 TATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATAT 4779

Qy    5806 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 5865
Db    4780 TACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCT 4839

Qy    5866 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 5925
Db    4840 GTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGA 4899

Qy    5926 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 5985
Db    4900 GATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCT 4959

Qy    5986 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 6045
Db    4960 GGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAG 5019

Qy    6046 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 6105
Db    5020 ATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATG 5079

Qy    6106 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 6160
Db    5080 TTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACTAC 5134
```

RESULT 3
US-09-738-877-1
; Sequence 1, Application US/09738877
; Patent No. US20020015970A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan
; APPLICANT: Weiss, Stephen J.
; APPLICANT: Glynne, Richard
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET
; TITLE OF INVENTION: SCREENING FOR ANGIOGENESIS MODULATORS
; FILE REFERENCE: A-69806/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/738,877
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,425
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT/US 00/22061
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 6608
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; AUTHORS: Nauert et al.
; TITLE: Gravin, an autoantigen recognized by serum from myastheinia gravis
; TITLE: patients, is a kinase scaffold protein
; JOURNAL: Curr. Biol.
; VOLUME: 7
; ISSUE: 1
; PAGES: 52-62
; DATE: 1997-01-01
; DATABASE ACCESSION NUMBER: U81607
; DATABASE ENTRY DATE: 1997-06-26
US-09-738-877-1

Query Match 34.8%; Score 2144.4; DB 10; Length 6608;
Best Local Similarity 64.3%; Pred. No. 0;
Matches 4215; Conservative 0; Mismatches 1796; Indels 543; Gaps 45;

```
Qy       3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db     162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221

Qy      63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db     222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281

Qy     120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db     282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341

Qy     171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db     342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401

Qy     231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG---------------------- 268
Db     402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461

Qy     269 --------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db     462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521

Qy     309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
Db     522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy     369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db     582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy     429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db     639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy     489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db     699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy     549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db     759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy     609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db     819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy     669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db     864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy     729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db     924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy     789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db     981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040
```

```
Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1461 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1496

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693

Qy   2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db   2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy   2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db   2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy   2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db   2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy   2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db   2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy   2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db   2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy   2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db   2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy   2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db   3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy   2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db   3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy   2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
```

```
Db   3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy   3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db   3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy   3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db   3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy   3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db   3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy   3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db   3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy   3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db   3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy   3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db   3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy   3360 GTCGCCACGTGCCA-----------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db   3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy   3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db   3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy   3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db   3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy   3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db   3774 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3833

Qy   3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db   3834 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3893

Qy   3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db   3894 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3953

Qy   3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db   3954 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4013

Qy   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db   4014 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4073

Qy   3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db   4074 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4133

Qy   3867 GAAAAGGAGAGTACTGAAGTGCAGAG---------------TCTTAGCCTGGAGGAGGGA 3911
Db   4134 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4193

Qy   3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db   4194 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4253

Qy   3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db   4254 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4313

Qy   4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db   4314 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4373

Qy   4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db   4374 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4433

Qy   4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db   4434 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4493

Qy   4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db   4494 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4553

Qy   4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db   4554 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4613

Qy   4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db   4614 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4673

Qy   4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db   4674 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4733

Qy   4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db   4734 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4784

Qy   4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db   4785 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4844

Qy   4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
Db   4845 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4904

Qy   4617 GCTTATGATTCACAGACCCAGGTTC---------------CTGCATGCAGGCTTGACAG 4660
Db   4905 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4964

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4965 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5024

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5025 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5084

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5085 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5144

Qy   4819 ------------------------------------------------------------ 4818
Db   5145 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5204

Qy   4819 ---------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5205 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5264

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5265 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5324

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5325 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5384
```

```
Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5385 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5444
Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5445 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5504
Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5505 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5564
Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5565 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5623
Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5624 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5682
Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5683 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5741
Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5742 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC------------------- 5781
Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5782 -----------------------------------------CATATAAATTCTGATTTA 5799
Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5800 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5859
Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5860 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5916
Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5917 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5976
Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5977 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6036
Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6037 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6096
Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6097 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6156
Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6157 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6216
Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6217 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6269
Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6270 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6329
Qy   5910 GTGCCTAAGGCCATGAGA----------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6330 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6389
Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6390 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6447
Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6448 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6507
Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6508 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6567
Qy   6134 CTGATTTAAATAAA 6147
Db   6568 CTGATTTAAATAAA 6581
```

RESULT 4
US-09-880-107-3439
; Sequence 3439, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3439
; LENGTH: 6608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U81607
US-09-880-107-3439

```
  Query Match             34.8%;  Score 2144.4;  DB 10;  Length 6608;
  Best Local Similarity   64.3%;  Pred. No. 0;
  Matches 4215;  Conservative    0;  Mismatches 1796;  Indels  543;  Gaps   45;

Qy      3 GGCCGGCGGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGG 62
Db    162 GGCTAGGCGCGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGC 221
Qy     63 AGCCCCGAGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCAT 119
Db    222 AGCCCGGAGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGC 281
Qy    120 GGGCCCGCAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGAC 170
Db    282 GGCCCCTCGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGAC 341
Qy    171 CCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAA 230
Db    342 CCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAG 401
Qy    231 CAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGG-------------------- 268
Db    402 CAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGT 461
Qy    269 -------------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGA 308
Db    462 CAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACGGAGGTTGGACAGAGA 521
Qy    309 GAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCT 368
```

```
Db    522 GACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTT 581

Qy    369 GTTGAAGATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCT 428
Db    582 GTTCACGACATCACAGATGATGGGCAGGAGGAGA---ACCGAAATATCGAACAGATTCCT 638

Qy    429 GCTTCAGAAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTT 488
Db    639 TCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATT 698

Qy    489 GGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAAT 548
Db    699 GGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACA 758

Qy    549 GAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAA 608
Db    759 GAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA 818

Qy    609 GCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCA 668
Db    819 ---------GGGGCTGGCGACCACCAGGACCCCAGCCTTGGGGCT------GGAGAAGCA 863

Qy    669 GCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAA 728
Db    864 GCATCCAAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGT 923

Qy    729 GAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAA 788
Db    924 GAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGT---GGAGGAA 980

Qy    789 GCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCG 848
Db    981 TGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCG 1040

Qy    849 AGCAGCCCAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGG 908
Db   1041 ACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGG 1100

Qy    909 GCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCC 968
Db   1101 GCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCA 1160

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAG 1028
Db   1161 GAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAG 1220

Qy   1029 CCAGCCTCGGAGGA--------------------------GCAGGAGCCGGCAGAAGAC 1061
Db   1221 GTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGT 1280

Qy   1062 ACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAG 1121
Db   1281 GCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAA 1340

Qy   1122 GTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTT 1181
Db   1341 GTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTT 1400

Qy   1182 GATGAGAAGATGGAAGCCCACCAAGAA---GTTGTTGCAGAGGTCCACGTGAGCACCGTG 1238
Db   1401 GATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTG 1460

Qy   1239 GAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAA 1298
Db   1461 GAGGAGAGAACCGAAGAGCAGAAA-----------------------ACGGAGGTGGAA 1496

Qy   1299 GGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAA 1358
Db   1497 GAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGGAATGGATGCAGAACCTCAGGAA 1556

Qy   1359 GCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCAC 1418
Db   1557 GCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCT 1616

Qy   1419 ACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATT 1478
Db   1617 ACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTT 1676

Qy   1479 GTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTG 1538
Db   1677 GTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTA 1736

Qy   1539 AAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAA 1598
Db   1737 AAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAA 1796

Qy   1599 CGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCA 1658
Db   1797 AGAGGA---GGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCG 1853

Qy   1659 GAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAG 1718
Db   1854 GACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAG 1913

Qy   1719 ACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGA 1778
Db   1914 ATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGA 1973

Qy   1779 ACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAG 1838
Db   1974 GCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAG 2033

Qy   1839 ATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTG 1898
Db   2034 ATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTG 2093

Qy   1899 GAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGAT 1958
Db   2094 GACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAA 2153

Qy   1959 GAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACT 2018
Db   2154 GAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACC 2213

Qy   2019 TCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCA 2078
Db   2214 TCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAAGG 2273

Qy   2079 TCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAG 2138
Db   2274 TCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGAT 2333

Qy   2139 GAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAG 2198
Db   2334 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 2393

Qy   2199 GACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGT 2258
Db   2394 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 2453

Qy   2259 GTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAA 2318
Db   2454 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 2513

Qy   2319 CTGGAAGAGAAAGCCGAAGAC-----------TCTAGTGTAGAGCAGTTGTCCACTGAG 2366
Db   2514 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 2573

Qy   2367 ATCGAACCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGG 2426
Db   2574 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 2633

Qy   2427 AAGAAAAGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAG 2486
Db   2634 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 2693
```

```
Qy  2487 ATAAATGAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTG 2546
Db  2694 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 2753

Qy  2547 GAGAGGGAGAAGATGG---------AAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTG 2597
Db  2754 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAGGCGCAGAGCAGCCCGAGCAGAAG 2813

Qy  2598 GGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCA 2657
Db  2814 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 2873

Qy  2658 GTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCC 2717
Db  2874 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 2933

Qy  2718 GCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAG 2777
Db  2934 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 2993

Qy  2778 GTCACTGAAAAAGACATCATTGC---AGAAGAAACTCCTGTGCTCACCCAGACGTTACCA 2834
Db  2994 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCA 3053

Qy  2835 GAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCT 2894
Db  3054 GAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCT 3113

Qy  2895 GTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGG 2954
Db  3114 GTGACAGCTGCAGAAACTGCAGGGCCATTGGGTTCCGAAGAAGGAACCGAAGCATCTGCT 3173

Qy  2955 GCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACC 3014
Db  3174 GCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACC 3233

Qy  3015 ACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAG 3074
Db  3234 ACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAA 3293

Qy  3075 GAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAG 3134
Db  3294 GAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAG 3353

Qy  3135 GTGCCTGCAACC---CAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAG 3191
Db  3354 CTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCAGAGGCAGAA 3413

Qy  3192 GAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAA 3251
Db  3414 AGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAA 3473

Qy  3252 GGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCT 3311
Db  3474 GTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACC 3533

Qy  3312 ACTCCAGAGAGCCTTG---AAGTTCCTGAAGTCACGGCAGATGTAGAC---------CAT 3359
Db  3534 ACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTT 3593

Qy  3360 GTCGCCACGTGCCA------------------GGTTATCAAGCTCCAGCAGCTGATGGAA 3401
Db  3594 GTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAA 3653

Qy  3402 CAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACT 3461
Db  3654 CAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACC 3713

Qy  3462 CCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAG 3521
Db  3714 CCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCAT 3773

Qy  3522 GACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACT 3581
Db  3774 GAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCT 3833

Qy  3582 GCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGG 3641
Db  3834 GCACAGAAAGAGAGGCCTCCAGCACCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAA 3893

Qy  3642 GAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCTGCAGCCGTG 3701
Db  3894 CAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTA 3953

Qy  3702 CCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGA---------GGTTGACTGGTTG 3752
Db  3954 TCCATTCTGTCAAAGACTGAGGGGACTCAAGAGGCTGACCAGTATGCTGATGAGAAAACC 4013

Qy  3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGG------ACCCAAC 3806
Db  4014 AAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGT 4073

Qy  3807 AGTCAAAAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAG 3866
Db  4074 CGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAA 4133

Qy  3867 GAAAAGGAGAGTACTGAAGTGCAGAG--------------TCTTAGCCTGGAGGAGGGA 3911
Db  4134 AAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGA 4193

Qy  3912 GAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAA 3971
Db  4194 GAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAA 4253

Qy  3972 GAAGG---TGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTC 4028
Db  4254 GAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTC 4313

Qy  4029 CTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGC 4088
Db  4314 CAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGC 4373

Qy  4089 CCTTCTC---TCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTG 4145
Db  4374 CCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCT 4433

Qy  4146 GACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAA------GGTCATAGAAACGGTT 4199
Db  4434 GAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCC 4493

Qy  4200 GTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAG 4259
Db  4494 AACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAG 4553

Qy  4260 AAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTG 4319
Db  4554 AAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACA 4613

Qy  4320 GGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACC 4379
Db  4614 GGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGC 4673

Qy  4380 CTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAG 4439
Db  4674 TTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTC 4733

Qy  4440 GACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTC 4499
Db  4734 GATGAGCAGGTTGCTTGCC------AGGAGGTCAAAGTGAGTGTAGCAATTG---AGGAT 4784

Qy  4500 CTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAAC 4559
Db  4785 TTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAAC 4844

Qy  4560 GTCATTCAGACAGCCGTTGACCAGTTCGCACGTAC---AGAAACAGCCCCCGAAACTCAT 4616
```

```
Db   4845 ATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTG 4904

Qy   4617 GCTTATGATTCACAGACCCAGGTTC----------------CTGCATGCAGGCTTGACAG 4660
Db   4905 ACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAG 4964

Qy   4661 CAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACACCCAGTGCC 4720
Db   4965 GAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACT 5024

Qy   4721 GCAGCCCA--------------------GAGAGGACTTGCAAGTCCTGACCGTTCTGGA 4759
Db   5025 TCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAA 5084

Qy   4760 GGCATGGGCTCAGCCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAA- 4818
Db   5085 GACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTGTAAAT 5144

Qy   4819 ------------------------------------------------------------ 4818
Db   5145 GATCAGCAGCTGGAAGAGGTCGTCCTCCCATCTGAGGAAGAGGGAGGTGGAGCTGGAACA 5204

Qy   4819 --------------------------------------AGTAAGCATTGAGAAGCTGCCT 4840
Db   5205 AAGTCTGTGCCAGAAGATGATGGTCATGCCTTGTTAGCAGAAAGAATAGAGAAGTCACTA 5264

Qy   4841 CCTCAACCCAAAGA--TCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTA 4898
Db   5265 GTTGAACCGAAAGAAGATGAAAAAGGTGATGATGTTGATGACCCTGAAAACCAGAACTCA 5324

Qy   4899 GCCCAGGCAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGA 4958
Db   5325 GCCCTGGCTGATACTGATGCCTCAGGAGGCTTAACCAAAGAGTCCCCAGATACAAATGGA 5384

Qy   4959 CCAAAGCTAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCT 5018
Db   5385 CCAAAACAAAAAGAGAAGGAGGATGCCCAGGAAGTAGAATTGCAGGAAGGAAAAGTGCAC 5444

Qy   5019 ACCAAGTCAGTCAAAGAGAACAAGGCCCAGGCAGAAGAGGACCTGCAGGAGCCAAAGGGA 5078
Db   5445 AGTGAATCAGATAAAGCGATCACACCCCAAGCACAGGAGGAGTTACAGAAACAAGAGAGA 5504

Qy   5079 GACCTGGCAGAATCCTAAGATGTTAGTTG--------------------CTCATTGTAC 5117
Db   5505 GAATCTGCAAAGTCAGAACTTACAGAATCTTAAAACATCATGCAGTTAAACTCATTGTCT 5564

Qy   5118 ATCTGTAAGACCAGAATGTGAAAACAAGTCACAGAACAA----GATGCTGCTGTTGGGAC 5173
Db   5565 GTTTGGAAGACCAGAATGTGAAGACAAGTAGTAGAAGAAAATGAATGCTGCTGCTGAGA- 5623

Qy   5174 CTTGAGACCAAGATTTCAGAGCCCATGAGATCCAGAGAGCAGGGCCGTCCAATGATTTCC 5233
Db   5624 CTGAAGACCAGTATTTCAGA-ACTTTGAGAATTGGAGAGCAGGCACATCAACTGATCTCA 5682

Qy   5234 ACCCAGTAGAGCACCCCGACAATTCTGAGGCTTCATCGGGAGCTAGAGCCAGCTAACATT 5293
Db   5683 TTTCTAGAGAGC-CCCTGACAATCCTGAGGCTTCATCAGGAGCTAGAGCCATTTAACATT 5741

Qy   5294 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCTAACATT 5353
Db   5742 TCCTCTTTCCAAGACCAACCTACAATTTTCCCTTGATAAC-------------------- 5781

Qy   5354 TCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTTGATGCCGTCCGTGTATTTCGGATTTA 5413
Db   5782 ------------------------------------------CATATAAATTCTGATTTA 5799

Qy   5414 AGGTCCTGCGTTCTCAACCTGGAAC-CAATTCTGCCATACCTAGTTCCACTTCTCAAACT 5472
Db   5800 AGGTCCTAAATTCTTAACCTGGAACTGGAGTTGGCAATACCTAGTTCTGCTTCTGAAACT 5859

Qy   5473 GGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTAT 5532
Db   5860 GGAGTATCATTCTTTACATATTTATATGTATGTTTTAAGTAG---TCCTCCTGTATCTAT 5916

Qy   5533 TGTATATTTTTTTCTAACGTTT-----------------AAGCACATGCTTTTTGTATTA 5575
Db   5917 TGTATATTTTTTTCTTAATGTTTAAGGAAATGTGCAGGATACTACATGCTTTTTGTATCA 5976

Qy   5576 TGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTG 5635
Db   5977 CACAGTATATGATGGGGCATGTGCCATAGTGCAGGCTTGGGGAGCTTTAAGCCTCAGTTA 6036

Qy   5636 TAACCTGCAGCAAACAGATAACATTC---------CTGGCAAGAAGAGACAAGTCTTTTT 5686
Db   6037 TATAACCCACAAAAAACAGAGCCTCCTAGATGTAACATTCCTGATCAAGGTACAATTCTT 6096

Qy   5687 TAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCT 5746
Db   6097 TAAAATTCACTAATGATTGAGGTCCATATTTAGTGGTACTCTGAAATTGGTCACTTTCCT 6156

Qy   5747 ATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATT 5806
Db   6157 ATTACACGGAGTGTGCCAAAACTAAAAAGCATTTTGAAACATACAGAATGTTCTATTGTC 6216

Qy   5807 ACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTG 5866
Db   6217 ATTGGGAAATTTTGC-------TTTCTAACCCAGTGGAGGTTAGAAAGAAGTTATATTCT 6269

Qy   5867 TGTCAGA-----------------TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGT 5909
Db   6270 GGTAGCAAATTAACTTTACATCCTTTTTCCTACTTGTTATGGTTGTTTGGACCGATAAGT 6329

Qy   5910 GTGCCTAAGGCCATGAGA---------------TGTTTCCTAATGCAGAAGGAATCTGT 5953
Db   6330 GTGCTTAATCCTGAGGCAAAGTAGTGAATATGTTTTATATGTTATGAAGAAAAGAATTGT 6389

Qy   5954 TGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTG 6013
Db   6390 TGTAAG--TTTTTGATTCTACTCTTATATGCTGGACTGCATTCACACATGGCATGAAATA 6447

Qy   6014 AGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTG 6073
Db   6448 AGTCAGGTTCTTTACAAATGGTATTTTGATAGATACTGGATTGTGTTTGTGCCATATTTG 6507

Qy   6074 TGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAA 6133
Db   6508 TGCCATTCCTTTAAGAACAATGTTGCAACACATTCATTTGGATAAGTTGTGATTTGACGA 6567

Qy   6134 CTGATTTAAATAAA 6147
Db   6568 CTGATTTAAATAAA 6581
```

RESULT 5
US-09-738-877-2
; Sequence 2, Application US/09738877
; Patent No. US20020015970A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan
; APPLICANT: Weiss, Stephen J.
; APPLICANT: Glynne, Richard
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND ME
; TITLE OF INVENTION: SCREENING FOR ANGIOGENESIS MODULATORS
; FILE REFERENCE: A-69806/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/738,877
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,425
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT/US 00/22061
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1

```
; SEQ ID NO 2
;   LENGTH: 5346
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-738-877-2

  Query Match             31.6%;  Score 1945.6;  DB 10;  Length 5346;
  Best Local Similarity   67.5%;  Pred. No. 0;
  Matches 3253;  Conservative    0;  Mismatches 1324;  Indels  243;  Gaps   27;

Qy         33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGC---GGGGAGCGAC 89
Db          1 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 60

Qy         90 ACGCCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGA------ 143
Db         61 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 120

Qy        144 ---GCAGCTGGAGACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGC 200
Db        121 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 180

Qy        201 CAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAAC 260
Db        181 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 240

Qy        261 CAGGAGGG------------------------------------------GCAGGAGGAA 278
Db        241 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 300

Qy        279 GAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGA 338
Db        301 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 360

Qy        339 GTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAG 398
Db        361 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 420

Qy        399 GAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAAATGGTA 458
Db        421 GAGA---ACCGAAATATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 477

Qy        459 CAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGT 518
Db        478 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 537

Qy        519 TTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACT 578
Db        538 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 597

Qy        579 GTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAG 638
Db        598 GTGAAGAAAGATGAAGGGGAGGGAGCAGCA---------GGGGCTGGCGACCACCAGGAC 648

Qy        639 CCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCC 698
Db        649 CCCAGCCTTGGGGCT------GGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 702

Qy        699 ACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAA 758
Db        703 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 762

Qy        759 GCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAG 818
Db        763 GCCGAATCTGGCCAA---GCAGTGGAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 819

Qy        819 AAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACAACATCT 878
Db        820 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 879

Qy        879 TCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAG 938
Db        880 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 939

Qy        939 AAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAA 998
Db        940 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 999

Qy        999 GTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGA---------------- 1042
Db       1000 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1059

Qy       1043 -----------GCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTAC 1091
Db       1060 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1119

Qy       1092 GAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAG 1151
Db       1120 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1179

Qy       1152 AAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAA--- 1208
Db       1180 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1239

Qy       1209 GTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGA 1268
Db       1240 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAGAAAAC---- 1295

Qy       1269 GGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAA 1328
Db       1296 --------------------GGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1335

Qy       1329 CTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGC 1388
Db       1336 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1395

Qy       1389 AGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTAAGTCCTGAAGAG 1448
Db       1396 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1455

Qy       1449 AAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAG 1508
Db       1456 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1515

Qy       1509 GAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAG 1568
Db       1516 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1575

Qy       1569 AAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGA 1628
Db       1576 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGAGGA---GGAGGAGACGAGGAATCAGGG 1632

Qy       1629 GAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGC 1688
Db       1633 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1692

Qy       1689 TCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAA 1748
Db       1693 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1752

Qy       1749 GCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAA 1808
Db       1753 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 1812

Qy       1809 GGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGA 1868
Db       1813 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 1872

Qy       1869 CCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCC 1928
Db       1873 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 1932

Qy       1929 ACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAG 1988
Db       1933 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 1992
```

```
Qy   1989 CCAGAGGAACCAAAGCGTAGGTGGATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTC 2048
Db   1993 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2052

Qy   2049 GGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGG 2108
Db   2053 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2112

Qy   2109 ACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACA 2168
Db   2113 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2172

Qy   2169 GACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAG 2228
Db   2173 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2232

Qy   2229 CCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTA 2288
Db   2233 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2292

Qy   2289 GTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAAGAC--------- 2339
Db   2293 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2352

Qy   2340 ---TCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTT 2396
Db   2353 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2412

Qy   2397 TCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGACGGGAAGCAAGAACAA 2456
Db   2413 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2472

Qy   2457 GCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTAATGTCCCAGCC 2516
Db   2473 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2532

Qy   2517 GTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGGGAGAAGATGG---------AAGCC 2567
Db   2533 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2592

Qy   2568 CAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGT 2627
Db   2593 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2652

Qy   2628 AAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGT 2687
Db   2653 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2712

Qy   2688 GTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACA 2747
Db   2713 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2772

Qy   2748 GCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGC---AGAA 2804
Db   2773 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 2832

Qy   2805 GAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTC 2864
Db   2833 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 2892

Qy   2865 ACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTC 2924
Db   2893 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 2952

Qy   2925 CGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCA 2984
Db   2953 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3012

Qy   2985 GTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTA 3044
Db   3013 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3072

Qy   3045 GAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCC 3104
Db   3073 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3132

Qy   3105 GTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACC---CAGACTGTGCAGAGA 3161
Db   3133 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3192

Qy   3162 ACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCT 3221
Db   3193 CTTCAGCCTGTGCAGAGAGCAGAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGT 3252

Qy   3222 TCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCAT 3281
Db   3253 CTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCT 3312

Qy   3282 CTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTG---AAGTTCCTGAA 3338
Db   3313 TTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAA 3372

Qy   3339 GTCACGGCAGATGTA---------GACCATGTCGCCACGTGCCA--------------- 3373
Db   3373 GTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCT 3432

Qy   3374 --GGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACC 3431
Db   3433 GGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACC 3492

Qy   3432 TTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGG 3491
Db   3493 CCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACA 3552

Qy   3492 ACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAG 3551
Db   3553 ACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAG 3612

Qy   3552 TCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCT 3611
Db   3613 TCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCACCTTCC 3672

Qy   3612 AATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCT 3671
Db   3673 AGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAGCAT 3732

Qy   3672 ACACAGCAAGAGCTTACTGCTGCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAA 3731
Db   3733 ACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACTCAA 3792

Qy   3732 GAGGGTGA---------GGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAACAGGAG 3782
Db   3793 GAGGCTGACCAGTATGCTGATGAGAAAACCAAAGACGTACCATTTTTCGAAGGACTTGAG 3852

Qy   3783 GTGTTTGTACACTCTGG------ACCCAACAGTCAAAAGGCTGCTGATGTGACATATGAC 3836
Db   3853 GGGTCTATAGACACAGGCATAACAGTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTTAAA 3912

Qy   3837 AGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAG---- 3892
Db   3913 GGTGAAGGGACAGAAGAAGCTGAATGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCAC 3972

Qy   3893 -----------TCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAA 3941
Db   3973 GCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAA 4032

Qy   3942 AGGGAGACAAAGCCAGAGCAAGTGAGTGAAG---AAGGTGAGCAGGAAACAGCCGCTCCT 3998
Db   4033 ACAGAAGCAGAGCCAACCCATGTGAATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACC 4092

Qy   3999 GAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGG 4058
Db   4093 GTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAGACAGTGAATGTGCCCATCATAGATGGG 4152

Qy   4059 GGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTC---TCCCAGACCAAGACAAAGCA 4115
```

```
Db    4153 GCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCA 4212

Qy    4116 GGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCT 4175

Db    4213 GTATGCACCAAAATTCAAGTTCAGAGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCA 4272

Qy    4176 GTGGAAAAGGTCAT------AGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAG 4229

Db    4273 GAGGAGGAAAAGGTCTTAGGAGAAACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAG 4332

Qy    4230 TGTGTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACT 4289

Db    4333 CCTGCAGGTGCACATTTAGTTCTGGAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCC 4392

Qy    4290 CTTCAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCA 4349

Db    4393 GCTCATCCAGGGGAAGATGCTGTGCCCACAGGGCCCGACTGTCAGGCAAAATCGACACCA 4452

Qy    4350 ATCATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGC 4409

Db    4453 GTGATAGTATCTGCTACTACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACC 4512

Qy    4410 GCATCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGAC 4469

Db    4513 ACATCACTGAAGTGGAAGTCAGATGAAGTCGATGAGCAGGTTGCTTGCC------AGGAG 4566

Qy    4470 GGCAAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTT 4529

Db    4567 GTCAAAGTGAGTGTAGCAATTG---AGGATTTAGAGCCTGAAAATGGGATTTTGGAACTT 4623

Qy    4530 GAGAGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCA 4589

Db    4624 GAGACCAAAAGCAGTAAACTTGTCCAAAACATCATCCAGACAGCCGTTGACCAGTTTGTA 4683

Qy    4590 CGTAC---AGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCA 4646

Db    4684 CGTACAGAAGAAACAGCCACCGAAATGTTGACGTCTGAGTTACAGACACAAGCTCACGTG 4743

Qy    4647 TGCAGGCTTGACAGCAGGGA 4666

Db    4744 ATAAAAGCTGACAGCCAGGA 4763
```

RESULT 6
US-09-917-800A-1223/c
; Sequence 1223, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1223
;   LENGTH: 563
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 AI233818
US-09-917-800A-1223

```
  Query Match             8.9%;  Score 550;  DB 10;  Length 563;
  Best Local Similarity   99.1%;  Pred. No. 1.9e-135;
  Matches  553;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy    5601 ATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAGATAACATT 5660

Db     563 AAAGCGACGCTTTGAAAAGCTCCAAGCTTCAACTGTAACCTGCAGCAAACAAATAACATT 504

Qy    5661 CCTGGCAAGAAGAGACAAGTCTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCT 5720

Db     503 CCTGGCAAGAAAAGACAAGTCTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCT 444

Qy    5721 AGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTT 5780

Db     443 AGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTT 384

Qy    5781 TGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTGCTAATCCAG 5840

Db     383 TGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTGCTAATCCAG 324

Qy    5841 TCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGA 5900

Db     323 TCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGA 264

Qy    5901 CAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGT 5960

Db     263 CAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGT 204

Qy    5961 TTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTG 6020

Db     203 TTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTG 144

Qy    6021 TTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCT 6080

Db     143 TTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCT 84

Qy    6081 TCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 6140

Db      83 TCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 24

Qy    6141 AAATAAACATATTTGACT 6158

Db      23 AAATAAACATATTTGCCT 6
```

RESULT 7
US-09-917-800A-1475
; Sequence 1475, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31

```
;  PRIOR APPLICATION NUMBER: US 60/222,040
;  PRIOR FILING DATE: 2000-07-31
;  PRIOR APPLICATION NUMBER: US 60/222,880
;  PRIOR FILING DATE: 2000-11-02
;  PRIOR APPLICATION NUMBER: US 60/290,029
;  PRIOR FILING DATE: 2001-05-11
;  PRIOR APPLICATION NUMBER: US 60/290,645
;  PRIOR FILING DATE: 2001-05-15
;  PRIOR APPLICATION NUMBER: US 60/292,336
;  PRIOR FILING DATE: 2001-05-22
;  PRIOR APPLICATION NUMBER: US 60/295,798
;  PRIOR FILING DATE: 2001-06-06
;  PRIOR APPLICATION NUMBER: US 60/297,457
;  PRIOR FILING DATE: 2001-06-13
;  PRIOR APPLICATION NUMBER: US 60/298,884
;  PRIOR FILING DATE: 2001-06-19
;  PRIOR APPLICATION NUMBER: US 60/303,459
;  PRIOR FILING DATE: 2001-07-09
;  NUMBER OF SEQ ID NOS: 1740
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1475
;   LENGTH: 178
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 U75404
US-09-917-800A-1475

  Query Match             2.6%;  Score 162.2;  DB 10;  Length 178;
  Best Local Similarity   97.8%;  Pred. No. 3.3e-33;
  Matches  175;  Conservative    0;  Mismatches    3;  Indels    1;  Gaps    1;

Qy     5962 TTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTGT 6021
Db        1 TTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTCACTCCTGT 60

Qy     6022 TCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTT 6081
Db       61 TCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGT-CCCCTT 119

Qy     6082 CTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 6140
Db      120 CTTTAAGAACAATGTTGCATTACGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 178

RESULT 8
US-10-101-487-74
; Sequence 74, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/101,487
;  CURRENT FILING DATE:  2002-03-20
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 74
;   LENGTH: 720
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: oligonucleotide
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2)..(718)
US-10-101-487-74

  Query Match             1.7%;  Score 106;  DB 9;  Length 720;
  Best Local Similarity   49.9%;  Pred. No. 6.5e-18;
  Matches  349;  Conservative    0;  Mismatches  345;  Indels    6;  Gaps    3;

Qy      946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db       15 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 74

Qy     1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db       75 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 134

Qy     1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db      135 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 194

Qy     1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCA---ACGGAAGTGTTTG 1182
Db      195 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 254

Qy     1183 ATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGA 1242
Db      255 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 314

Qy     1243 AGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAA 1302
Db      315 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 374

Qy     1303 CAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTG 1362
Db      375 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 434

Qy     1363 AGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTC 1422
Db      435 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 494

Qy     1423 AACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCA 1482
Db      495 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 554

Qy     1483 GTGAGGTGGAGATGCTGTCCTCTCAGGAA-AGAATCAAGGTACAGGGAAGTCCCTTGAAG 1541
Db      555 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 614

Qy     1542 AAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACG- 1600
Db      615 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 674

Qy     1601 -AGGAGGTGGGGAGACGAAGAGCCTGGAGAATACCAACA 1639
Db      675 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGA 714

RESULT 9
US-10-101-487-76/c
; Sequence 76, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/101,487
;  CURRENT FILING DATE:  2002-03-20
```

```
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 76
;   LENGTH: 720
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: oligonucleotide
US-10-101-487-76

  Query Match             1.7%;  Score 106;  DB 9;  Length 720;
  Best Local Similarity   49.9%;  Pred. No. 6.5e-18;
  Matches  349;  Conservative    0;  Mismatches  345;  Indels    6;  Gaps    3;

Qy    946 AAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACG 1005
Db    710 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 651

Qy   1006 AGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAG 1065
Db    650 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 591

Qy   1066 ACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTG 1125
Db    590 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 531

Qy   1126 GTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCA---ACGGAAGTGTTTG 1182
Db    530 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 471

Qy   1183 ATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGA 1242
Db    470 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 411

Qy   1243 AGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAA 1302
Db    410 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 351

Qy   1303 CAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTG 1362
Db    350 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 291

Qy   1363 AGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTC 1422
Db    290 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 231

Qy   1423 AACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCA 1482
Db    230 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 171

Qy   1483 GTGAGGTGGAGATGCTGTCCTCTCAGGAA-AGAATCAAGGTACAGGGAAGTCCCTTGAAG 1541
Db    170 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGG 111

Qy   1542 AAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACG- 1600
Db    110 AAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGAGGAAGAGG 51

Qy   1601 -AGGAGGTGGGGAGACGAAGAGCCTGGAGAATACCAACA 1639
Db     50 AGGAAGAGGAGGAAGAGGAAGAGGAGGAAGAGGAGGAAGA 11

RESULT 10
US-10-101-487-71
; Sequence 71, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/101,487
;  CURRENT FILING DATE:  2002-03-20
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 71
;   LENGTH: 522
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: oligonucleotide
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(522)
US-10-101-487-71

  Query Match             1.5%;  Score 89.6;  DB 9;  Length 522;
  Best Local Similarity   50.2%;  Pred. No. 1.1e-13;
  Matches  221;  Conservative    0;  Mismatches  219;  Indels    0;  Gaps    0;

Qy    937 AGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAA 996
Db      5 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 64

Qy    997 AAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAG 1056
Db     65 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 124

Qy   1057 AAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAG 1116
Db    125 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 184

Qy   1117 ACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAG 1176
Db    185 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 244

Qy   1177 TGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCG 1236
Db    245 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 304

Qy   1237 TGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAG 1296
Db    305 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 364

Qy   1297 AAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGG 1356
Db    365 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 424

Qy   1357 AAGCTGAGCCTGCTGAGGAG 1376
Db    425 AGGAAGAGGAAGAAGAGGAG 444

RESULT 11
US-10-101-487-73/c
; Sequence 73, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
```

```
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/101,487
;  CURRENT FILING DATE:  2002-03-20
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 73
;   LENGTH: 530
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: oligonucleotide
US-10-101-487-73

  Query Match             1.5%;  Score 89.6;  DB 9;  Length 530;
  Best Local Similarity   50.2%;  Pred. No. 1.1e-13;
  Matches  221;  Conservative    0;  Mismatches  219;  Indels    0;  Gaps    0;

Qy      937 AGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAA 996
Db      526 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 467

Qy      997 AAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAG 1056
Db      466 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 407

Qy     1057 AAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAG 1116
Db      406 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 347

Qy     1117 ACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAG 1176
Db      346 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 287

Qy     1177 TGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCG 1236
Db      286 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 227

Qy     1237 TGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAG 1296
Db      226 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 167

Qy     1297 AAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGG 1356
Db      166 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 107

Qy     1357 AAGCTGAGCCTGCTGAGGAG 1376
Db      106 AGGAAGAGGAAGAAGAGGAG 87

RESULT 12
US-10-101-487-69
; Sequence 69, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/101,487
;  CURRENT FILING DATE:  2002-03-20
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 69
;   LENGTH: 554
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: oligonucleotide
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(528)
US-10-101-487-69

  Query Match             1.5%;  Score 89.6;  DB 9;  Length 554;
  Best Local Similarity   50.2%;  Pred. No. 1.2e-13;
  Matches  221;  Conservative    0;  Mismatches  219;  Indels    0;  Gaps    0;

Qy      937 AGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAA 996
Db       17 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 76

Qy      997 AAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAG 1056
Db       77 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 136

Qy     1057 AAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAG 1116
Db      137 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 196

Qy     1117 ACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAG 1176
Db      197 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 256

Qy     1177 TGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCG 1236
Db      257 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 316

Qy     1237 TGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAG 1296
Db      317 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 376

Qy     1297 AAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGG 1356
Db      377 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 436

Qy     1357 AAGCTGAGCCTGCTGAGGAG 1376
Db      437 AGGAAGAGGAAGAAGAGGAG 456

RESULT 13
US-10-101-487-106
; Sequence 106, Application US/10101487
; Patent No. US20020169125A1
; GENERAL INFORMATION:
;  APPLICANT: LEUNG, DAVID W.
;  APPLICANT:  BERGMAN, PHILIP A.
;  APPLICANT:  LOFQUIST, ALAN
;  APPLICANT:  PIETZ, GREGORY E.
;  APPLICANT:  TOMPKINS, CHRISTOPHER K.
;  APPLICANT:  WAGGONER JR., DAVID W.
;  TITLE OF INVENTION: RECOMBINANT PRODUCTION OF POLYANIONIC POLYMERS AND USES
;  TITLE OF INVENTION:  THEREOF
;  FILE REFERENCE: 077319/0329
;  CURRENT APPLICATION NUMBER: US/10/101,487
;  CURRENT FILING DATE:  2002-03-20
;  PRIOR APPLICATION NUMBER: 60/277,705
;  PRIOR FILING DATE: 2001-03-21
;  NUMBER OF SEQ ID NOS: 116
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 106
;   LENGTH: 554
```

```
;   TYPE: DNA
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence: Synthetic
;   OTHER INFORMATION: oligonucleotide
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(537)
US-10-101-487-106

  Query Match             1.5%;  Score 89.6;  DB 9;  Length 554;
  Best Local Similarity   50.2%;  Pred. No. 1.2e-13;
  Matches  221;  Conservative    0;  Mismatches  219;  Indels    0;  Gaps    0;

Qy     937 AGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAA 996
           || ||     |||| || ||    |||    |   ||   ||   | | ||| | || | ||||
Db      17 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 76

Qy     997 AAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAG 1056
           | | ||| || ||||| |||||   ||||| |   ||||  ||     || ||| ||||    | ||
Db      77 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 136

Qy    1057 AAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAG 1116
           | ||     ||  | |    |      || |||  | |||  | |  |||   |        ||||
Db     137 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 196

Qy    1117 ACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAG 1176
           |  | |   |  ||   ||||| |     |||||| ||    |              || |    ||||
Db     197 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 256

Qy    1177 TGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCG 1236
            |    || ||  |    ||| |    |   |||| |   |  | ||| |     | |            |
Db     257 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 316

Qy    1237 TGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAG 1296
            |||  ||   ||| |||||| | | || ||    | ||||  |||| ||   |   |   | ||
Db     317 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 376

Qy    1297 AAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGG 1356
           | | |    | ||||     |        || |     |  ||       ||||||       | |
Db     377 AGGAAGAGGAAGAAGAGGAGGAAGAAGAAGAGGAAGAGGAAGAAGAGGAGGAAGAAGAAG 436

Qy    1357 AAGCTGAGCCTGCTGAGGAG 1376
           | |  |||    |   ||||||
Db     437 AGGAAGAGGAAGAAGAGGAG 456


RESULT 14
US-09-864-761-20733
; Sequence 20733, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 20733
;   LENGTH: 575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005822.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.4
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  12
;   OTHER INFORMATION: NT HIT: AF185255.1, EVALUE 4.80e+00
;   OTHER INFORMATION: EST_HUMAN HIT: BE071749.1, EVALUE 5.80e+00
US-09-864-761-20733

  Query Match             1.2%;  Score 71;  DB 10;  Length 575;
  Best Local Similarity   53.3%;  Pred. No. 1e-08;
  Matches  172;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy     969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAA-GAAAAGGAAAAGACAGA 1027
           ||| ||   | ||||| | || |   ||   ||| ||| ||| ||  || ||   || |  ||| |||
Db     212 GAGGAGGAAGAGGAGGAGGAAGAGGAGGAAGAAGAGGAGAAAGGAGGAGAAGAAGGAAGA 271

Qy    1028 GCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGA 1087
           |          |||||| |||||   |  || ||   |||  |||    |   |    |  ||
Db     272 GGATCAGGAAGAGGAGGAGGAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAGGA 331

Qy    1088 CTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGA 1147
              | ||||||  |||   |      |||||| |||  |  ||   ||| |      |  |||
Db     332 AGAGGAGAAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGA 391

Qy    1148 GGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGA 1207
           |||| ||    |           || || ||| |     || ||| ||    || |  || |   |
Db     392 GGAGGAGGAAGAGAAGAAGGAAAAGGAGGAAGAGGAGGAGGAGGAAGAGGAGCAGAAGAA 451

Qy    1208 AGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGG 1267
             |   |    ||||   | |            | | || ||| ||| |||||| ||| ||  |
Db     452 GGAAGAGGAGGAGGAAGAGGAAGAAGAGGAGAAGGAGAAAGAAGAGGAGGAGGAAGAGGA 511

Qy    1268 AGGAGAGGCTGAAGGGGGCGTGG 1290
           ||  ||||  ||||  ||   | |
Db     512 AGAGGAGGAGGAAGCTGGACTTG 534
```

```
RESULT 15
US-09-864-761-3972
; Sequence 3972, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
;  APPLICANT: Penn, Sharron G.
;  APPLICANT:  Rank, David R.
;  APPLICANT:  Hanzel, David K.
;  APPLICANT:  Chen, Wensheng
;  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
;  TITLE OF INVENTION:  GENE EXPRESSION ANALYSIS BY MICROARRAY
;  FILE REFERENCE: Aeomica-X-1
;  CURRENT APPLICATION NUMBER: US/09/864,761
;  CURRENT FILING DATE:  2001-05-23
;  PRIOR APPLICATION NUMBER: US 60/180,312
;  PRIOR FILING DATE: 2000-02-04
;  PRIOR APPLICATION NUMBER: US 60/207,456
;  PRIOR FILING DATE: 2000-05-26
;  PRIOR APPLICATION NUMBER: US 09/632,366
;  PRIOR FILING DATE: 2000-08-03
;  PRIOR APPLICATION NUMBER: GB 24263.6
;  PRIOR FILING DATE: 2000-10-04
;  PRIOR APPLICATION NUMBER: US 60/236,359
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: PCT/US01/00666
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00667
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00664
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00669
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00665
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00668
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00663
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00662
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00661
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: PCT/US01/00670
;  PRIOR FILING DATE: 2001-01-30
;  PRIOR APPLICATION NUMBER: US 60/234,687
;  PRIOR FILING DATE: 2000-09-21
;  PRIOR APPLICATION NUMBER: US 09/608,408
;  PRIOR FILING DATE: 2000-06-30
;  PRIOR APPLICATION NUMBER: US 09/774,203
;  PRIOR FILING DATE: 2001-01-29
;  NUMBER OF SEQ ID NOS: 49117
;  SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 3972
;   LENGTH: 1969
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005822.1
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL =  13
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL =  11
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =  14
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 9.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL =  12
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 9.4
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 11
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL =  12
US-09-864-761-3972

  Query Match             1.2%;  Score 71;  DB 10;  Length 1969;
  Best Local Similarity   53.3%;  Pred. No. 2.4e-08;
  Matches  172;  Conservative    0;  Mismatches  150;  Indels    1;  Gaps    1;

Qy    969 GAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAA-GAAAAGGAAAAGACAGA 1027

Db    554 GAGGAGGAAGAGGAGGAGGAAGAGGAGGAAGAAGAGGAGAAAGGAGGAGAAGAAGGAAGA 613

Qy   1028 GCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGA 1087

Db    614 GGATCAGGAAGAGGAGGAGGAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAGGA 673

Qy   1088 CTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGA 1147

Db    674 AGAGGAGAAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGAGGAGGAGGAAGAGAAGGAAGA 733

Qy   1148 GGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGA 1207

Db    734 GGAGGAGGAAGAGAAGAAGGAAAAGGAGGAAGAGGAGGAGGAGGAAGAGGAGCAGAAGAA 793

Qy   1208 AGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGG 1267

Db    794 GGAAGAGGAGGAGGAAGAGGAAGAAGAGGAGGAAGGAGAAAGAAGAGGAGGAGGAAGAGGA 853

Qy   1268 AGGAGAGGCTGAAGGGGGCGTGG 1290

Db    854 AGAGGAGGAGGAAGCTGGACTTG 876


Search completed: December 13, 2002, 00:10:46
Job time : 308 secs
```

```
OM nucleic - nucleic search, using sw model

Run on:          December 12, 2002, 14:13:30 ; Search time 7323 Seconds
                                              (without alignments)
                                              13623.405 Million cell updates/sec

Title:           US-09-902-432-3
Perfect score:   6160
Sequence:        1 cgggccggcgggagtagaag..........aaataaacatatttgactac 6160

Scoring table:   IDENTITY_NUC
                 Gapop 10.0 , Gapext 1.0

Searched:        16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters:      32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       EST:*
                 1:  em_estba:*
                 2:  em_esthum:*
                 3:  em_estin:*
                 4:  em_estmu:*
                 5:  em_estov:*
                 6:  em_estpl:*
                 7:  em_estro:*
                 8:  em_htc:*
                 9:  gb_est1:*
                 10:  gb_est2:*
                 11:  gb_htc:*
                 12:  gb_est3:*
                 13:  gb_est4:*
                 14:  gb_est5:*
                 15:  em_estfun:*
                 16:  em_estom:*
                 17:  gb_gss:*
                 18:  em_gss_hum:*
                 19:  em_gss_inv:*
                 20:  em_gss_pln:*
                 21:  em_gss_vrt:*
                 22:  em_gss_fun:*
                 23:  em_gss_mam:*
                 24:  em_gss_mus:*
                 25:  em_gss_other:*
                 26:  em_gss_pro:*
                 27:  em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	751.4	12.2	1153	14	D88456	D88456 D88456 Mous
c	2	682.8	11.1	710	14	BQ200192	BQ200192 UI-R-DQ1-
c	3	634.8	10.3	683	13	BI274459	BI274459 UI-R-CW0-
	4	571	9.3	776	13	BG974884	BG974884 602842771
	5	552.8	9.0	1269	11	BC015830	BC015830 Homo sapi
c	6	550.4	8.9	746	13	BM230239	BM230239 K0292D03-
c	7	550	8.9	563	9	AI233818	AI233818 EST230506
	8	547.6	8.9	661	13	BI652823	BI652823 603298753
c	9	532.2	8.6	710	13	BM119923	BM119923 L0932E09-
	10	529.2	8.6	668	10	BB546856	BB546856 BB546856
c	11	512	8.3	703	14	BQ176546	BQ176546 UI-M-DJ2-
	12	501.8	8.1	610	9	AI225639	AI225639 uj13g09.y
c	13	492.2	8.0	665	13	BM228830	BM228830 K0268G02-
c	14	467.8	7.6	476	9	AI103489	AI103489 EST212778
	15	460.2	7.5	573	12	BG670384	BG670384 DRNBAE11
	16	459.6	7.5	679	14	BQ443215	BQ443215 UI-M-EV0-
	17	458.6	7.4	567	12	BF286252	BF286252 EST450843
c	18	456.6	7.4	623	13	BM226613	BM226613 K0230F11-
c	19	453.8	7.4	625	13	BM226974	BM226974 K0236B12-
	20	450.6	7.3	530	9	AI156819	AI156819 ui44c01.y
c	21	442.8	7.2	601	13	BM226680	BM226680 K0231G01-
c	22	442.8	7.2	614	13	BM245338	BM245338 K0723D03-
	23	439.6	7.1	754	13	BG966814	BG966814 602834309
	24	437.8	7.1	745	12	BG663107	BG663107 DRA08D03
	25	436.6	7.1	534	10	AW919250	AW919250 EST350554
c	26	434.8	7.1	593	13	BM225990	BM225990 K0220D09-
c	27	430.8	7.0	586	13	BM293486	BM293486 C0305H04-
c	28	429.4	7.0	453	10	AW535062	AW535062 UI-R-BS0-
	29	421.4	6.8	641	10	BB665546	BB665546 BB665546
c	30	418.2	6.8	577	13	BM230475	BM230475 K0295H08-
c	31	416.6	6.8	428	9	AI176113	AI176113 EST219692
c	32	415.8	6.8	574	13	BM230613	BM230613 K0298D07-
c	33	414.8	6.7	574	13	BM229218	BM229218 K0275C05-
c	34	412	6.7	589	13	BM244908	BM244908 K0717D04-
	35	410.2	6.7	886	12	BF163549	BF163549 601769945
	36	408.6	6.6	644	13	BI446672	BI446672 dac81f11.
	37	406.6	6.6	674	9	AL045541	AL045541 DKFZp434K
	38	405.8	6.6	704	13	BI645683	BI645683 603275189
c	39	400.4	6.5	562	14	BQ554344	BQ554344 H4027E12-
c	40	400.2	6.5	483	9	AA963479	AA963479 UI-R-E1-g
c	41	392.6	6.4	552	13	BM124130	BM124130 L0535F11-
c	42	385.2	6.3	599	9	AI317366	AI317366 uj13g09.x
c	43	384.2	6.2	550	13	BM244784	BM244784 K0715G12-
	44	382	6.2	599	14	BQ566352	BQ566352 gi58b03.y
c	45	371.6	6.0	522	13	BM197966	BM197966 C0352A03-

ALIGNMENTS

```
RESULT 1
D88456
LOCUS       D88456                  1153 bp    mRNA    linear   EST 01-APR-1998
DEFINITION  D88456 Mouse ICR testis (J.Tsuchida) Mus musculus cDNA clone A-12,
            mRNA sequence.
ACCESSION   D88456
VERSION     D88456.1  GI:1669580
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1153)
  AUTHORS   Taketo,M.M., Araki,Y., Matsunaga,A., Yokoi,A., Tsuchida,J., Nishina
            ,Y., Nozaki,M., Tanaka,H., Koga,M., Uchida,K., Matsumiya, Okuyama
            ,A., Rochelle,J.M., Nishimune,Y., Matsui,M. and Seldin,M.
  TITLE     Mapping of eight testis-specific genes to mouse chromosomes
  JOURNAL   Genomics 46 (1), 138-142 (1997)
  MEDLINE   98066773
COMMENT     Contact: Junji Tsuchida
            Department of Science for Laboratory Animal Experimentation
            Research Institute for Microbial Diseases, Osaka University
            3-1 Yamadaoka, Suita, Osaka 565, Japan
            Tel: 06-879-8338
            Fax: 06-879-8339
            Email: tsuchida@biken.osaka-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .1153
                     /organism="Mus musculus"
```

```
                     /strain="ICR"
                     /db_xref="taxon:10090"
                     /clone="A-12"
                     /clone_lib="Mouse ICR testis (J.Tsuchida)"
                     /sex="male"
                     /tissue_type="testis"
BASE COUNT      367 a    285 c    299 g    201 t      1 others
ORIGIN

  Query Match             12.2%;  Score 751.4;  DB 14;  Length 1153;
  Best Local Similarity   81.1%;  Pred. No. 1.6e-159;
  Matches  975;  Conservative    0;  Mismatches  157;  Indels   70;  Gaps    6;

Qy       4081 GAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTTCAAA 4140
                | |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db          1 GGGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGATTGCATAGAGGTTCAAGTTCAAA 60

Qy       4141 GCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACGGTTG 4200
              ||    ||||| ||||||||||||| | |||||||| ||||||||  |||||| || |
Db         61 GCTCAGACACACCAGTCACTCAAACAACCGAAGCTGTGAAAAAGGTCGAAGAAACTGTGG 120

Qy       4201 TGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCTGAGA 4260
                 |||||||   | |||||||     |||||| ||||| || | |||||||||||||
Db        121 CAACTTCAGAGATGGATGAAAGTTTGGAGTGTGCAGGTGCGCAATCATTACCAGCTGAGA 180

Qy       4261 AGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCCCTGG 4320
              ||  ||| | ||| ||||| || |||||||||||||| |||||||| || |||| ||
Db        181 AGCTCTCCGAAACCGGTGGCTACGGGACTCTTCAGCATGGAGAGGACACCGTGCCCCAGG 240

Qy       4321 GGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGCACCC 4380
               |||||||||||||||| |||||||||| ||||||||||||||||||||||| ||
Db        241 NGCCTGAGTCTCAGGCAGAGTCCATCCCAATAATAGTAACTCCTGCTCCTGAAAGCATCC 300

Qy       4381 TACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAAGAGG 4440
              ||||| ||||||| ||| ||||| | ||||||||||||| | || ||||||| |||||
Db        301 TACATTCTGACCTTCAAAGAGAAGTGAGCGCATCCCAGAAACAGAGATCAGATGAAGATA 360

Qy       4441 ACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAAGTCC 4500
              ||||||||||||||||||||||||||||| |||||||||||| ||||||   || ||| |||
Db        361 ACAAGCCAGATGCTGGTCCTGATGCTGCCGGCAAGGAGAGTGCAGCAAGAGAGAAAATCC 420

Qy       4501 TCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTGAACG 4560
              ||| ||||||||||||||| |||||||||||||||||||| |||||||| | ||  |
Db        421 TCAGGGCTGAACCTGAGATCTTGGAACTTGAGAGTAAGAGCAATAAGATTGTCCAGAGTG 480

Qy       4561 TCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCATGCTT 4620
              |||| |||||||||| |||||||| ||||||||||||||||||||||||| || ||||
Db        481 TCATCCAGACAGCCGTCGACCAGTTTGCACGTACAGAAACAGCCCCCGAAACCCACGCTT 540

Qy       4621 ATGATTCACAGACCCAGGTTCCTGCATGCAGGCTTGACAGCAGGGAGCCCAACAGATGCT 4680
               ||||| |||||  |||||||||       |  ||     |||||||| |||||||||||
Db        541 CTGATTTACAGAATCAGGTTCCT-------GTGATGCAGCCAGGGAGCACAACAGATGCT 593

Qy       4681 GGACAAAAATGAAAG---------------ATGCCAAGATGAAACACCCAGTGCCGCAGC 4725
              ||||||| |||||||                |||||||| ||||| ||||||| |||
Db        594 GGACAAAGATGAAAGCGACCTTCAAGTCTCCCCCCAAGATGGAACACTCAGTGCCGTAGC 653

Qy       4726 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAGCCTCGGAAATGCT 4785
              ||||  |||||||||   | ||||  |||||| ||||||| ||    || || |||||| |
Db        654 CCAGGAAGGACTTGCGGTTTCTGATAGTTCTGAAGGCATGAGCAAGGCTTCAGAAATGAT 713

Qy       4786 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAGTAAGCATTGAGAAGCTGCCTCCTCA 4845
                 || |||||||||||||||| ||| ||||||||| |||  ||||||||||||||| |||
Db        714 CACCACGCTTGCAGTTGAAAGTGCCAGTGTCAAAGAAAGTGTTGAGAAGCTGCCTCTTCA 773

Qy       4846 ACCCAAAGATCCAAAAGGAGCATGCTGCTGATGGCCCTCAGCTCCAAAGCTTAGCCCAGG 4905
                 |||||||  ||||||| |||||||||||| ||||| ||||  ||||||||||||| |||
Db        774 GTGCAAAGAT-GAAAAGGAACATGCTGCTGACGGCCCCCAGCACCAAAGCTTAGCCAAGG 832

Qy       4906 CAGAGGCCAGTGCCTCTGGAAACCTAACCAAAGAATCCCCAGACACCAACGGACCAAAGC 4965
              ||||||    ||||||||||| |||||||||| ||||||||||||||||||||||||||
Db        833 CAGAGGCGGATGCCTCTGGAAATCTAACCAAAGAGTCCCCAGACACCAACGGACCAAAGC 892

Qy       4966 TAACCGAGGAGGGCGATCCCCCAAAAGTTGAGGTCCAGGAAGAAGAAATGTCTACCAAGT 5025
              ||||||||| || ||| |||  ||||  |
Db        893 TAACCGAGGAAGGAGATGCCCTGAAAGAAG------------------------------ 922

Qy       5026 CAGTCAAAGAGAACAAGGCCCAGGCAGAAGAG---GACCTGCAGGAGCCAAAGGGAGACC 5082
                    ||  ||||||||||| |||||||      ||||| |||||||||||||||||
Db        923 ------AAATGAACAAGGCCCAGACAGAAGAGGACGACCTACAGGAGCCAAAGGGAGACC 976

Qy       5083 TGGCAGAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAAC 5142
              || ||||||||||||| || ||||||||||        ||||| || |||||||||
Db        977 TGACAGAATCCTAAGACGTGAGTTGCTCAT--------TGTAAAACTAGAATGTGAAGTG 1028

Qy       5143 AAGTCACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAG 5202
              |||||| ||||||||||||||||||||||||||||||| ||||||||||||| ||||
Db       1029 AAGTCACGGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAAATTTCAGAGCCCTTGAG 1088

Qy       5203 ATCCAGAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAG 5262
               | |||||||||| ||||||||||||||| |||| | ||||||||||||||| |||||
Db       1089 GTGCAGAGAGCAGAGCCGTCCAATGATTTCAACCCTGCAGAGCACCCCGACAATCCTGAG 1148

Qy       5263 GC 5264
              ||
Db       1149 GC 1150


RESULT 2
BQ200192/c
LOCUS       BQ200192                 710 bp    mRNA    linear   EST 02-MAY-2002
DEFINITION  UI-R-DQ1-clu-e-23-0-UI.s1 UI-R-DQ1 Rattus norvegicus cDNA clone
            UI-R-DQ1-clu-e-23-0-UI 3', mRNA sequence.
ACCESSION   BQ200192
VERSION     BQ200192.1  GI:20416657
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized rat cell line R3327-5a library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..710
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-DQ1-clu-e-23-0-UI"
                     /clone_lib="UI-R-DQ1"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DQ1
                     library is a normalized Rat Cell Line R3327-5A library
```

```
                  (nR5A) constructed in pT3T7 PAC vector according to the
                  procedure described by Bonaldo, Lennon & Soares
                  (Normalization and Subtraction:  Two Approaches to
                  Facilitate Gene Discovery.  Genome Research 6:  791-806,
                  1996).  The oligonucleotide used to prime first strand
                  synthesis contained the sequence tag GGACTAGATC between
                  the Not I cloning site and dT18 stretch. The Rat Cell Line
                  R3327-5A was provided by Mary Hendrix of the University of
                  Iowa.
                  TAG_LIB=UI-R-DQ1
                  TAG_TISSUE=rat cell line R3327-5a
                  TAG_SEQ=GGACTAGATC"
BASE COUNT      246 a    131 c    134 g    198 t      1 others
ORIGIN

  Query Match             11.1%;  Score 682.8;  DB 14;  Length 710;
  Best Local Similarity   99.6%;  Pred. No. 5.6e-144;
  Matches  684;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy       5472 TGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTA 5531
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        710 TGGAGCATCCTCCTTTATGTATTTATATGTATGTTGTATGTAGTCCTCCTCCTGTACCTA 651

Qy       5532 TTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACGGG 5591
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        650 TTGTATATTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACGGG 591

Qy       5592 TGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACA 5651
              |||||||||||||||||||| |||||||||||||||||||||||||||||||||||||||
Db        590 TGTGCAGCCATAGCGACGCTNTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACA 531

Qy       5652 GATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTG 5711
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        530 GATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTG 471

Qy       5712 TGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAA 5771
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        470 TGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAA 411

Qy       5772 ACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTG 5831
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        410 ACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTG 351

Qy       5832 CTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAA 5891
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        350 CTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAA 291

Qy       5892 TGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCT 5951
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        290 TGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCT 231

Qy       5952 GTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAG 6011
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        230 GTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAG 171

Qy       6012 TGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATC 6071
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        170 TGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATC 111

Qy       6072 TGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGAC 6131
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        110 TGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGAC 51

Qy       6132 AACTGATTTAAATAAACATATTTGACT 6158
              |||||||||||||||||||||||| ||
Db         50 AACTGATTTAAATAAACATATTTGCCT 24


RESULT 3
BI274459/c
LOCUS       BI274459                 683 bp    mRNA    linear   EST 18-JUL-2001
DEFINITION  UI-R-CW0-bwj-b-05-0-UI.s1 UI-R-CW0 Rattus norvegicus cDNA clone
            UI-R-CW0-bwj-b-05-0-UI 3', mRNA sequence.
ACCESSION   BI274459
VERSION     BI274459.1  GI:14885393
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 683)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized rat aorta pool library cDNA Library Preparation: M.B.
            Soares Lab Clone distribution: clones will be available through
            Research Genetics (www.resgen.com)
            Seq primer: M13 Forward
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..683
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CW0-bwj-b-05-0-UI"
                     /clone_lib="UI-R-CW0"
                     /dev_stage="ADULT"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CW0
                     library is a normalized library constructed from the
                     following rat aorta tissues: embryonic day 19, embryonic
                     day 21, adult day 1, adult day 12, adult day 75, adult day
                     200. For a detailed description of the library from which
                     this clone was derived, please visit our web site at
                     ratest.eng.uiowa.edu. The subtraction has been previously
                     described in (Bonaldo, Lennon and Soares, Genome Research
                     6:791-806, 1996)
                     TAG_LIB=UI-R-CW0
                     TAG_TISSUE=rat aorta pool
                     TAG_SEQ=CTGTAGGATC"
BASE COUNT      235 a    125 c    129 g    194 t
ORIGIN

  Query Match             10.3%;  Score 634.8;  DB 13;  Length 683;
  Best Local Similarity   99.4%;  Pred. No. 4.3e-133;
  Matches  658;  Conservative    0;  Mismatches    2;  Indels    2;  Gaps    2;

Qy       5497 TATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAA 5556
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db        683 TATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATTG-ATATTTTTTTCTAACGTTTAA 625

Qy       5557 GCACATGCTTTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAA 5616
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        624 GCACATGCTTTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAA 565

Qy       5617 AAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAGATAACATTCCTGGCAAGAAGAGAC 5676
              ||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        564 AAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAAATAACATTCCTGGCAAGAAGAGAC 505

Qy       5677 AAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGG 5736
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       504 AAGTCTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGG 445

Qy      5737 TTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATG 5796
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       444 -TGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATG 386

Qy      5797 TCCCAATATTACCATGATTTTTCCCCCCTTTTGCTAATCCAGTCCAGGTTGGAAAGAA 5856
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       385 TCCCAATATTACCATGATTTTTCCCCCCTTTTGCTAATCCAGTCCAGGTTGGAAAGAA 326

Qy      5857 GTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTA 5916
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       325 GTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTA 266

Qy      5917 AGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTC 5976
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       265 AGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTC 206

Qy      5977 TTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTA 6036
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       205 TTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTA 146

Qy      6037 TTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGT 6096
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       145 TTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGT 86

Qy      6097 TGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGA 6156
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        85 TGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGC 26

Qy      6157 CT 6158
             ||
Db        25 CT 24
```

RESULT 4
BG974884

```
LOCUS       BG974884                 776 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602842771F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4978744 5',
            mRNA sequence.
ACCESSION   BG974884
VERSION     BG974884.1  GI:14362521
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
            Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10975  row: m  column: 17
            High quality sequence stop: 753.
FEATURES             Location/Qualifiers
     source          1..776
                     /organism="Mus musculus"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4978744"
                     /clone_lib="NCI_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and  Differentiation 7, 3-11 (1996)."
BASE COUNT      278 a    181 c    216 g    101 t
ORIGIN

  Query Match            9.3%;  Score 571;  DB 13;  Length 776;
  Best Local Similarity  86.9%;  Pred. No. 1.3e-118;
  Matches  677;  Conservative    0;  Mismatches   90;  Indels   12;  Gaps    4;

Qy       327 GAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAG 386
             ||||||||||||| | |||||||||| ||||  |||||||  ||||||||||||||||||
Db         1 GAAAAAGACCGAGCTAAAGAAATGGCAGCCAGTTCCACAGTGGTTGAAGATATCACAAAG 60

Qy       387 GATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTG 446
             || | ||||||||| ||| | |||||||| |||||||||||||||||||| | |||||||
Db        61 GACGAGCAGGAGGAAACACCGGAAATAATCGAACAGATCCCTGCTTCAGAGAGCAATGTG 120

Qy       447 GAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTT 506
             ||||||||||  ||| ||||||||||||| |||||||| || ||||||||||| ||||||
Db       121 GAAGAAATGGCGCAGGCTGCTGAGTCCCAAGCTAATGACGTCGGCTTCAAGAAGGTATTT 180

Qy       507 AAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTC 566
             ||||||||||||||||||||||||||||||||||||||||| |||||||||||||| |||
Db       181 AAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAACGAAAAGTCAGATACCGTC 240

Qy       567 CAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGA 626
             || |||||||||||||||||||||||||||||||||||||||||||| |||||||| |||
Db       241 CAGCTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCCGTCGGAGCAGGA 300

Qy       627 GACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAG 686
             |||||||| |||||| | |||||||   ||||||| || |||||||||||||||||||||
Db       301 GACCACCAAGAGCCCGGAGTGGAGA---CCGTCGGCGAATCAGCATCCAAAGAAAGTGAG 357

Qy       687 CTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAA 746
             ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db       358 CTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGCACAGAGCAGCACAGAA 417

Qy       747 ATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAA 806
             || ||||||||||||||||||| ||||| | | |||||||||||     | || ||||||
Db       418 ATTCCCCTTCAAGCCGAATCTGGTCAAGGGACCGAGGAAGAAGCAGCCAAAGATGGAGAA 477

Qy       807 GAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGT 866
             ||||| | ||||||||| || |||||| ||| |||||| |||| |||||| |||| || |
Db       478 GAAAACCGAGAGAAAGAACCTACCAAGCCCCTAGAATCTCCGACCAGCCCTGTCAGCAAT 537

Qy       867 GAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAG 926
             ||||||||||||||||||||||| |||||||||||||| |||||||||||||||||||||
Db       538 GAGACAACATCTTCCTTCAAGAAATTCTTCACTCACGGCTGGGCCGGCTGGCGCAAGAAG 597

Qy       927 ACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAA 986
             ||||||||||||||| |||| || ||||||||||||||| ||||||||||| ||||||||
Db       598 ACCAGCTTCAAGAAACCAAAGGAAGATGATCTGGAAACTTCCGAGAAGAGACAGGAGCAA 657

Qy       987 GAGGCAGAAAAAGTAGACGAGGAAGAAAAGG---AAAAGACAGAGCCAGCCTCGGAGGAG 1043
             ||||| ||||||||||||   | |   ||||     ||||||||||||||| | |  |||
Db       658 GAGGCTGAAAAAGTAGACCGAGGAAGCAAGGGGACCAAGACAGAGCCAGCCCCAGCCGAG 717

Qy      1044 CAGGAG-----CCGGCAGAAGACACAGACCAGGCCAGGTTG-TCAGCAGACTACGAGAA 1096
              || ||     | |    | | |||||||||||||||| ||  |||| ||||| |||||
Db       718 GAGCAGGAAGCCTGCAGAAGGGCACAGACCAGGCCAGGATGACCAGCCGACTATGAGAA 776
```

RESULT 5
BC015830

```
LOCUS       BC015830                1269 bp    mRNA    linear   HTC 17-OCT-2001
DEFINITION  Homo sapiens, clone IMAGE:4691862, mRNA.
ACCESSION   BC015830
VERSION     BC015830.1  GI:16198369
```

```
KEYWORDS    HTC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1269)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 38 Row: k Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4826749
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..1269
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4691862"
                     /tissue_type="Lung"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
BASE COUNT      419 a    311 c    365 g    174 t
ORIGIN

  Query Match            9.0%;  Score 552.8;  DB 11;  Length 1269;
  Best Local Similarity  69.8%;  Pred. No. 1.8e-114;
  Matches  894;  Conservative    0;  Mismatches  287;  Indels   99;  Gaps    7;

Qy        10 GGGAGTAGAAGAGCCACTGAGCCATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCG 69
Db         2 GGGGGGAGAAGTGCGGAGGAGCCATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGG 61

Qy        70 AGCAGCCG---GCGGGGAGCGACACGCCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCG 126
Db        62 AGCAGCCGCCCGAGGGGAGCTCCACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCT 121

Qy       127 CAGCTGAAGCCTCGGGA---------GCAGCTGGAGACCCCGCCGACGCGGACCCCGCCA 177
Db       122 CGGCCGAGGCGGCGCCAGACACCACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCA 181

Qy       178 CCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCTGTCAACGGCGTAGCTGAACAAGGAG 237
Db       182 CCAAGCTCCTACAGAAGAATGGTCAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATG 241

Qy       238 ATGTCCATGTCCAAGAGGAAAACCAGGAGGG----------------------------- 268
Db       242 AGCTCAGCCTCCAGGAGGGTGACCTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAG 301

Qy       269 -------------GCAGGAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAG 315
Db       302 CCCTAAACAGCCAGGAGGAAGAAGAAGTCATTGTCACAGAGGTTGGACAGAGAGACTCTG 361

Qy       316 AAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAG 375
Db       362 AAGATGTGAGCAAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACG 421

Qy       376 ATATCACAAAGGATGGGCAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAG 435
Db       422 ACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCTTCTTCAG 481

Qy       436 AAAACAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCA 495
Db       482 AAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATTGGATTTA 541

Qy       496 AGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGT 555
Db       542 AGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGC 601

Qy       556 CAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTG 615
Db       602 CTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCAGCA------- 654

Qy       616 TCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCA 675
Db       655 --GGGGCTGGCGACCACAAGGACCCCAGCCTTGGGGCT------GGAGAAGCAGCATCCA 706

Qy       676 AAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGA 735
Db       707 AAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAA 766

Qy       736 GCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAG 795
Db       767 GCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAA---GCAGTGGAGGAATGCAAAG 823

Qy       796 ATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCC 855
Db       824 AGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTC 883

Qy       856 CAGTCAACAGTGAGACAACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCT 915
Db       884 CCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCT 943

Qy       916 GGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGA 975
Db       944 GGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGA 1003

Qy       976 GAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCT 1035
Db      1004 AAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCT 1063

Qy      1036 CGGAGGA--------------------------GCAGGAGCCGGCAGAAGACACAGACC 1068
Db      1064 CCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACG 1123

Qy      1069 AGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTG 1128
Db      1124 AGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTG 1183

Qy      1129 ACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGA 1188
Db      1184 GCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAAA 1243

Qy      1189 AGATGGAAGCCCACCAAGAA 1208
Db      1244 AAAAAAAAAAAAAAAAAAAA 1263


RESULT 6
BM230239/c
LOCUS       BM230239                 746 bp    mRNA    linear   EST 31-JAN-2002
DEFINITION  K0292D03-3 NIA Mouse Unfertilized Egg cDNA Library (Long) Mus
            musculus cDNA clone K0292D03 3', mRNA sequence.
ACCESSION   BM230239
VERSION     BM230239.1  GI:17793419
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 746)
  AUTHORS   Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Stagg,C.A.,
            Martin,P., Aiba,K., Tanaka,T. and Ko,M.S.H.
  TITLE     Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library
            (Long)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: K0292  row: D  column: 03
            Seq primer: -21M13 Forward
            High quality sequence stop: 746
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..746
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:K0292D03-3"
                     /db_xref="taxon:10090"
                     /clone="K0292D03"
                     /clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long
                     )"
                     /tissue_type="Unfertilized Egg"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     extracted from a pool of 1488 unfertilized eggs.
                     Double-stranded cDNAs were synthesized with an Oligo(dT)
                     primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'],
                     treated with T4 DNA polymerase, and purified by
                     ethanol-precipitation. The cDNAs were ligated to
                     Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                     The DH10B E. coli host was transformed with the ligation
                     mixture by the standard chemical method. The average
                     insert size is about 2.5 kb. The library was constructed
                     by Yulan Piao (NIA)."
BASE COUNT      261 a    137 c    147 g    201 t
ORIGIN

  Query Match             8.9%;  Score 550.4;  DB 13;  Length 746;
  Best Local Similarity   90.4%;  Pred. No. 6e-114;
  Matches  678;  Conservative    0;  Mismatches   61;  Indels   11;  Gaps    8;

Qy       5412 TAAGGTCCTGCGTTCTCAACCTGGAACCAAT-TCTGCCATACCTAGTTCCACTTCTCAAA 5470
              |||||||||| ||||| ||||||||||||| || || ||||||||| |||||||||||||
Db        746 TAAGGTCCTGCTTTCTCCACCTGGAACCAATGTCGGCAATACCTAGTCCCACTTCTCAAA 687

Qy       5471 CTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCT 5530
              |||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        686 CTGGAGCCTCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCT 627

Qy       5531 ATTGTATA-TTTTTTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACG 5589
              |||||||| |||||||||| ||||||||||||||| |||||||||||||| ||||||| |
Db        626 ATTGTATATTTTTTTTCTAAGGTTTAAGCACATGCCTTTTGTATTATGCAGTATATAATG 567

Qy       5590 GGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAA 5649
              ||||||| ||||||||| |||||| |||| ||||||||||||||||||||||||||||||
Db        566 GGTGTGCAACCATAGCGAAGCTTTGAGAAGCCCCAAGCCTCAACTGTAACCTGCAGCAAA 507

Qy       5650 CAG----ATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTA 5705
              |||    |||||||||||||| ||||| ||||||| |||||||||||||||||||||||
Db        506 CAGAGAAATAACATTCCTGGCAGGAAGATACAAGTC-TTTTTAAAGTTTACTGATGCTTA 448

Qy       5706 GATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAA 5765
               ||||||||  |||||||||||||||||||||||||||||||||||||| |||||||||
Db        447 ACTCTGTGGGCTCGTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGTGAGCTCAGAA 388

Qy       5766 ATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCC 5825
              |||||| ||||||||| |||| ||| ||||||||| |  |  ||||
Db        387 ATAAAAACTCCATTTTGAGACATTCAGAATGTCCCAACATTACGACAACATTTT-TTTTT 329

Qy       5826 TTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTC 5885
              ||||| |||||||||||||||||||||||||||||  ||||||||||||||||   ||
Db        328 TTTTTTCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTTAGTGTCAGATTAAGCCCCATC 269

Qy       5886 TCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAG 5945
              |||||  ||||||||| |||||||||| | ||||||||| | |||||||| |||||||
Db        268 TCTTAACAATATGGACAGATGAGTGTGCC-ATGGCCATGAGCTATTTCCTAACGCAGAAG 210

Qy       5946 GAATCTGTTGTACGTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGA 6005
              |||| ||||||    ||||| |||||||||||||||  |||||||||||||||||||||
Db        209 GAATTTGTTGT-TACTTTTTGGATTGTACTCTTCTAGGCTGGACCGAATTCATATGCAGA 151

Qy       6006 TCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGT 6065
               |||||||| |||| ||||||||||||||||||||||||||||||||||||||||||| |
Db        150 TTGAAGTGAGACCTGCTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTAT 91

Qy       6066 TATATCTGTGCCCCTTC-TTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTG 6124
              ||||||||||||||||| ||||||||||||||| ||||||||||||||||||||||||||
Db         90 TATATCTGTGCCCCTTCTTTTAAGAACAATGTTACATTATGTTCCTTTGGATAAATTGTG 31

Qy       6125 ATTTGACAACTGATTTAAATAAACATATTT 6154
              |||||||||||||||| |||||| ||||||
Db         30 ATTTGACAACTGATTTCAATAAAAAATATTT 1


RESULT 7
AI233818/c
LOCUS       AI233818                 563 bp    mRNA    linear   EST 31-JAN-1999
DEFINITION  EST230506 Normalized rat lung, Bento Soares Rattus sp. cDNA clone
            RLUCR55 3' end, mRNA sequence.
ACCESSION   AI233818
VERSION     AI233818.1  GI:3817698
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 563)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     Other_ESTs: TC58449
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..563
                     /organism="Rattus sp."
                     /db_xref="ATCC (inhost):2039638"
                     /db_xref="taxon:10118"
                     /clone="RLUCR55"
                     /clone_lib="Normalized rat lung, Bento Soares"
                     /note="Organ: lung; Vector: pT7T3Pac; Site_1: EcoRI;
                     Site_2: NotI"
```

```
BASE COUNT      191 a     107 c     110 g     155 t
ORIGIN

  Query Match              8.9%;  Score 550;  DB 9;  Length 563;
  Best Local Similarity   99.1%;  Pred. No. 7.2e-114;
  Matches  553;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy       5601 ATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAGATAACATT 5660
Db        563 AAAGCGACGCTTTGAAAAGCTCCAAGCTTCAACTGTAACCTGCAGCAAACAAATAACATT 504

Qy       5661 CCTGGCAAGAAGAGACAAGTCTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCT 5720
Db        503 CCTGGCAAGAAAAGACAAGTCTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCT 444

Qy       5721 AGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTT 5780
Db        443 AGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTT 384

Qy       5781 TGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAG 5840
Db        383 TGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAG 324

Qy       5841 TCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGA 5900
Db        323 TCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGA 264

Qy       5901 CAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGT 5960
Db        263 CAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGT 204

Qy       5961 TTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTG 6020
Db        203 TTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTG 144

Qy       6021 TTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCT 6080
Db        143 TTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCT 84

Qy       6081 TCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 6140
Db         83 TCTTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTT 24

Qy       6141 AAATAAACATATTTGACT 6158
Db         23 AAATAAACATATTTGCCT 6


RESULT 8
BI652823
LOCUS       BI652823                 661 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603298753F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5339244 5',
            mRNA sequence.
ACCESSION   BI652823
VERSION     BI652823.1  GI:15567059
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11861  row: j  column: 13
            High quality sequence stop: 661.
FEATURES             Location/Qualifiers
     source          1..661
                     /organism="Mus musculus"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5339244"
                     /clone_lib="NCI_CGAP_Mam3"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
BASE COUNT      183 a     179 c     201 g      98 t
ORIGIN

  Query Match              8.9%;  Score 547.6;  DB 13;  Length 661;
  Best Local Similarity   90.2%;  Pred. No. 2.6e-113;
  Matches  597;  Conservative    0;  Mismatches   64;  Indels    1;  Gaps    1;

Qy       2656 CAGTCATTGATGGACCAGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATAT 2715
Db          1 CGGTCATTGATGGACCA-GGCAGTCACCAGTGCCGAAGAGCGGTCCCCTTCGTGGATAT 59

Qy       2716 CCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAG 2775
Db         60 CTGCTTCCATGACAGAACCTCTTGAGCACGCAGAGGGAGTGGCCACACCGCCTGTTGGAG 119

Qy       2776 AGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAG 2835
Db        120 AGGTCACTGAAAAAGACATCACTGCAGAAGCAACTCCTGCACTCGCCCAGACTTTACCAG 179

Qy       2836 AGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTG 2895
Db        180 GGGGCAAAGATGCCCATGACGACATAGTCACCAGTGAAGTGGATTTTACCTCAGAAGCAG 239

Qy       2896 TGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGG 2955
Db        240 TGACAGCCGCAGAAACCACAGAGGCGCTCCGCGCTGAAGAACTTACCGAAGCATCAGGGG 299

Qy       2956 CCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCA 3015
Db        300 CAGAAGAGACCACAGACATGGTGTCTGCAGTTTCCCAGCTGTCCGACTCCCCGGACACCA 359

Qy       3016 CAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGG 3075
Db        360 CAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGGGTGGCATGCTAGATACGGAAGAACAGG 419

Qy       3076 AGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGG 3135
Db        420 AGCGCCAGACGCAGGCCGTCCTCCAAGCCGTTGCAGACAAAGTGAAAGAGGACTCCCAGG 479

Qy       3136 TGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGG 3195
Db        480 TGCCTGCAACCCAGACTCTGCAGAGAGCAGGGCCGAAAGCACTGGAGAAGGTGGAGGAGG 539

Qy       3196 TAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGAC 3255
Db        540 TAGAGGAGGACTCCGAGGTGCTGGCTACCGAGAAAGAGAAGGATGTTGTGCCGGAAGGAC 599

Qy       3256 CCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTC 3315
Db        600 CCGTGCAGGAAGCTGAAACTGAGCATCTTGCACAGGGCTCCGAGACTGTACAGGCTACCC 659

Qy       3316 CA 3317
Db        660 CA 661
```

```
RESULT 9
BM119923/c
LOCUS       BM119923                 710 bp    mRNA    linear   EST 01-FEB-2002
DEFINITION  L0932E09-3 NIA Mouse Newborn Kidney cDNA Library (Long) Mus
            musculus cDNA clone L0932E09 3', mRNA sequence.
ACCESSION   BM119923
VERSION     BM119923.1  GI:17087949
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 710)
  AUTHORS   Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Pantano,S., Lim,M.K.
            and Ko,M.S.H.
  TITLE     Systematic Analyses of NIA Mouse Newborn Kidney cDNA Library
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: L0932  row: E  column: 09
            Seq primer: -21M13 Forward
            High quality sequence stop: 710
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..710
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:L0932E09-3"
                     /db_xref="taxon:10090"
                     /clone="L0932E09"
                     /clone_lib="NIA Mouse Newborn Kidney cDNA Library (Long)"
                     /tissue_type="Newborn Kidney"
                     /dev_stage="Newborn"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pSPORT1 (Invitrogen);
                     Site_1: SalI; Site_2: NotI; Mouse cDNA project by the
                     Laboratory of Genetics, National Institute on Aging (NIA),
                     Intramural Research Program, NIH
                     (http://lgsun.grc.nia.nih.gov/cDNA). This is a
                     long-transcript enriched cDNA library (Ref. Genome Res.
                     11:1553-1558 (2001). [PMID:11544199]). In brief
                     ,double-stranded cDNAs were synthesized with an Oligo(dT)
                     primer [Invitrogen: 5'-
                     pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTTT-3'] from 26
                     microgram of total RNA, treated with T4 DNA polymerase,
                     and purified by ethanol-precipitation. The cDNAs were
                     ligated to Lone-linker LL-Sal4, purified by
                     phenol/chloroform, and separated from free linkers by
                     Centricon 100. Then, the cDNAs were amplified by
                     long-range high fidelity PCR using Ex Taq polymerase
                     (Takara) with a primer Sal4-S. The products were purified
                     by phenol/chloroform and Centricon 100. The cDNAs were
                     digested with SalI and NotI enzymes, and cloned into
                     SalI/NotI site of pSPORT1 plasmid vector. The DH10B E.
                     coli host was transformed with ligation mixture by the
                     standard chemical method. The average insert size is about
                     3.0 kb. The library was constructed by Yulan Piao (NIA)."
BASE COUNT      252 a    130 c    137 g    191 t
ORIGIN

  Query Match             8.6%;  Score 532.2;  DB 13;  Length 710;
  Best Local Similarity   90.5%;  Pred. No. 8e-110;
  Matches  636;  Conservative    0;  Mismatches   58;  Indels    9;  Gaps    6;

Qy       5459 CCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCT 5518
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        710 CCACTTCTCAAACTGGAGCCTCCTCCTTTATGTATTTATATGTATGTTTTATGTAGTCCT 651

Qy       5519 CCTCCTGTACCTATTGTATA-TTTTTTTCTAACGTTTAAGCACATGCTTTTGTATTATG 5577
              ||||||||||||||||||| ||||||||||| ||||||||||||| |||||||||||
Db        650 CCTCCTGTACCTATTGTATATTTTTTTCTAAGGTTTAAGCACATGCCTTTTGTATTATG 591

Qy       5578 CAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTA 5637
              || ||||||| ||||||||||||||||| ||||||| |||| ||||||||||||||
Db        590 CAGTATATAATGGGTGTGCAGCCATAGCGAAGCTTTGAGAAGCCCCAAGCCTCAACTGTA 531

Qy       5638 ACCTGCAGCAAACAG----ATAACATTCCTGGCAAGAAGAGACAAGTCTTTTTAAAGTT 5693
              ||||||||||||||     |||||||||||||| ||||| ||||||| ||||||||||
Db        530 ACCTGCAGCAAACAGAGAAATAACATTCCTGGCAGGAAGATACAAGTC-TTTTAAAGTT 472

Qy       5694 TACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACA 5753
              ||||||||| ||  |||||||||  |||||||||||||||||||||||||||||||||
Db        471 TACTGATGCCTAACTCTGTGGGCTCGTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACA 412

Qy       5754 GCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTACCATGA 5813
              | ||||||||||||||||| ||||||||| |||| ||| |||||||| ||||| |  |
Db        411 GTGAGCTCAGAAATAAAAACTCCATTTTGAGACATTCAGAATGTCCCAACATTACGACAA 352

Qy       5814 TTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGA 5873
              ||||      ||| | |||||||||||||||||||||||||||||  ||||||||
Db        351 CATTTTTTTTTTTTTCTTCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTTAGTGTCAGA 292

Qy       5874 TTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGATGTTTC 5933
              ||||||||  ||||||||  ||||||||| ||||||||||| | ||||||||| | ||||
Db        291 TTAAGCCCCATCTCTTAACAATATGGACAGATGAGTGTGCC-ATGGCCATGAGCTATTTC 233

Qy       5934 CTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGGACCGAA 5993
              |||| |||||||||| |||||      ||||| |||||||||||||| ||||||||||
Db        232 CTAACGCAGAAGGAATTTGTTGT-TACTTTTTGGATTGTACTCTTCTAGGCTGGACCGAA 174

Qy       5994 TTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGATACTGGA 6053
              |||||||||||| ||||||| |||| |||||||||||||||||||||||||||||||
Db        173 TTCATATGCAGATTGAAGTGAGACCTGCTCTTTACAGATGGTATTTTGATAGATACTGGA 114

Qy       6054 GTTTGTCTGTGTTATATCTGTGCCCCTTC-TTTAAGAACAATGTTGCATTATGTTCCTTT 6112
              ||||||||| ||||||||||||||||| |||||||||||||| |||||||||||||
Db        113 GTTTGTCTGTATTATATCTGTGCCCCTTCTTTTAAGAACAATGTTACATTATGTTCCTTT 54

Qy       6113 GGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTG 6155
              ||||||||||||||||||||||||||| |||||| |||||||
Db         53 GGATAAATTGTGATTTGACAACTGATTTCAATAAAAATATTTG 11


RESULT 10
BB546856
LOCUS       BB546856                 668 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB546856 RIKEN full-length enriched, 0 day neonate eyeball Mus
            musculus cDNA clone E130314I19 3' similar to AB020886 Mus musculus
            mRNA for SSeCKS, mRNA sequence.
ACCESSION   BB546856
VERSION     BB546856.2  GI:16447501
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 31, 2000 this sequence version replaced gi:9618284.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
```

```
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..668
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="E130314I19"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     eyeball"
                     /tissue_type="eyeball"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      199 a    146 c    229 g     93 t      1 others
ORIGIN

  Query Match             8.6%;  Score 529.2;  DB 10;  Length 668;
  Best Local Similarity   88.8%;  Pred. No. 3.8e-109;
  Matches  596;  Conservative    0;  Mismatches   69;  Indels    6;  Gaps    2;

Qy       1183 ATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGA 1242
              |||||||| |||  |||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGAGAAGACGGAGNCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGA 60

Qy       1243 AGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAA 1302
              |||    | | | |||  ||||||| |||||  |||||||  |||||||| ||||||
Db         61 AGATGACGAAAGGGCA---AGGAGGAGCAGAGGTGGAAGGGGATGTGGTGGTGGAAGGAT 117

Qy       1303 CAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTG 1362
              | ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db        118 CGGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGACCCAGGAGGTCCCCCAGGAAGCTG 177

Qy       1363 AGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTC 1422
              |||||  |||||||||||||||| || |||   | || |||||||| || ||||| ||||
Db        178 AGCCTGTGGAGGAGCTGATGAAGACCAAAGAAGTATGCGTCTCTGGGGGTGACCATACTC 237

Qy       1423 AACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCA 1482
              | |||||||| |||||||||||||||||| ||| |||||||||| |||||||||||
Db        238 AGCTGACAGATCTAAGTCCTGAAGAGAAGATGCTACCCAAACACCCCGAAGGCATTGTCA 297

Qy       1483 GTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGA 1542
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||| ||||||
Db        298 GTGAGGTGGAGATGCTGTCCTCTCAGGAGAGAATCAAGGTACAGGGAAGTCCCCTGAAGA 357

Qy       1543 AACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAA---AC 1599
              | ||||||| || || |||||||||||||| || ||||||||||||||||||||    |
Db        358 AGCTCTTCAGCAGTTCGGGCTTAAAGAAGCTCTCCGGGAAGAAGCAGAAGGGGAAGAGAG 417

Qy       1600 GAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAG 1659
              |||||| ||||||| |||||||| |||||||||||||||||||| ||||| ||||||
Db        418 GAGGAGGCGGGGGAGATGAAGAGCCAGGAGAATACCAACACATTCAAACCGAGTCCCCAG 477

Qy       1660 AGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGA 1719
              ||||||||| |||||||||||||||||||||| || ||||| || ||||| ||||||
Db        478 AGAGTGCTGACGAGCAGAAGGGAGAGAGCTCTGCCTCTTCCCCTGAAGAGCCCGAGGAGA 537

Qy       1720 CCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAA 1779
               | |||||||||||| |||||   |||||||||||||| | |||||||||||||||
Db        538 TCGCGTGTCTGGAGAAGGGGCCATCGGAAGCACCCCAGGAAGCGGAAGCTGAGGAAGGAG 597

Qy       1780 CTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGA 1839
              | ||||||| |||||||||| || |||||||||| ||||||||||||||||||||    |
Db        598 CGACTTCCGACGGAGAGAAGAAAAGGGAAGGGATCACCCCCTGGGCATCCTTCAAAAAGA 657

Qy       1840 TGGTGACACCC 1850
              |||||||||||
Db        658 TGGTGACACCC 668


RESULT 11
BQ176546/c
LOCUS       BQ176546                 703 bp    mRNA    linear   EST 30-APR-2002
DEFINITION  UI-M-DJ2-bvw-p-22-0-UI.s1 NIH_BMAP_DJ2 Mus musculus cDNA clone
            UI-M-DJ2-bvw-p-22-0-UI 3', mRNA sequence.
ACCESSION   BQ176546
VERSION     BQ176546.1  GI:20352038
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 703)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Chin, H
            National Institute of Mental Health
            6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
            20892-9643, USA
            Tel: 301 443 1706
            Fax: 301 443 9890
            Email: mEST@mail.nih.gov
            Tissue Procurement: Dr. Robin Davisson
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
             The following repetitive elements were found in this cDNA
```

```
            sequence: 1-22, >AT_rich#Low_complexity (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..703
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="UI-M-DJ2-bvw-p-22-0-UI"
                     /clone_lib="NIH_BMAP_DJ2"
                     /tissue_type="subfornical organ and postrema"
                     /dev_stage="Adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /note="Organ: brain; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-M-DJ2 is a subtracted cDNA library constructed
                     according to Bonaldo, Lennon and Soares, Genome Research,
                     6:791-806, 1996. First strand cDNA synthesis was primed
                     with an oligo-dT primer containing a Not I site. Double
                     stranded cDNA was ligated to an EcoR I adaptor, digested
                     with Not I, and cloned directionally into pT7T3-Pac
                     vector. The oligonucleotide used to prime the synthesis of
                     first-strand cDNA contains a library tag sequence that is
                     located between the Not I site and the (dT)18 tail. The
                     sequence tag for this library is: GCTACATGAT, subfornical
                     organ and area postrema.
                     TAG_LIB=UI-M-DJ2
                     TAG_TISSUE=subfornical organ and postrema
                     TAG_SEQ=GCTACATGAT"
BASE COUNT      249 a    127 c    124 g    203 t
ORIGIN

  Query Match             8.3%;  Score 512;  DB 14;  Length 703;
  Best Local Similarity   90.6%;  Pred. No. 3.1e-105;
  Matches  614;  Conservative    0;  Mismatches   55;  Indels    9;  Gaps    6;

Qy       5484 CTTTATGTATTTATATGTATGTTTATGTAGTCCTCCTCCTGTACCTATTGTATA-TTTT 5542
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        703 CTTTATGTATTTATATGTATGTTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTT 644

Qy       5543 TTTCTAACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACGGGTGTGCAGCCAT 5602
              |||||| ||||||||||||| |||||||||||||||| ||||||| ||||||||||||||
Db        643 TTTCTAAGGTTTAAGCACATGCCTTTTGTATTATGCAGTATATAATGGGTGTGCAGCCAT 584

Qy       5603 AGCGACGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAG----ATAACA 5658
              ||||| ||||||| |||| ||||||||||||||||||||||||||||||     ||||||
Db        583 AGCGAAGCTTTGAGAAGCCCCAAGCCTCAACTGTAACCTGCAGCAAACAGAGAAATAACA 524

Qy       5659 TTCCTGGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTT 5718
              |||||||| ||||| ||||||| |||||||||||||||||||||||  ||||||||||
Db        523 TTCCTGGCAGGAAGATACAAGTC-TTTTTAAAGTTTACTGATGCTTAACTCTGTGGGCTC 465

Qy       5719 CTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCAT 5778
               |||||||||||||||||||||||||||||||||| ||||||||||||||||| ||||
Db        464 GTAGTCCTCTGAAAGTGGTTGTTTTCCTATGCACAGTGAGCTCAGAAATAAAAACTCCAT 405

Qy       5779 TTTGAAACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCC 5838
              |||| |||| ||| ||||||||| ||||| |  ||||        ||||| ||||||||
Db        404 TTTGAGACATTCAGAATGTCCCAACATTACGACAACATTTTTTTTTTTTTTTTCTAATCC 345

Qy       5839 AGTCCAGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATG 5898
              ||||||||||||||||||||||| ||||||||||||||| |||||||| |||||
Db        344 AGTCCAGGTTGGAAAGAAGTCTCCTTAGTGTCAGATTAAGCCCCATCTCTTAACAATATG 285

Qy       5899 GACAAATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTAC 5958
              |||| |||||||||| | ||||||||| | |||||||| |||||||||| ||||||
Db        284 GACAGATGAGTGTGCC-ATGGCCATGAGCTATTTCCTAACGCAGAAGGAATTTGTTGT-T 227

Qy       5959 GTTTTTTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCC 6018
               ||||| |||||||||||||||| |||||||||||||||||||||||| |||||||| ||
Db        226 ACTTTTTGGATTGTACTCTTCTAGGCTGGACCGAATTCATATGCAGATTGAAGTGAGACC 167

Qy       6019 TGTTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCC 6078
              || ||||||||||||||||||||||||||||||||||||||||||| |||||||||||||
Db        166 TGCTCTTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTATTATATCTGTGCCC 107

Qy       6079 CTTC-TTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGA 6137
              |||| ||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db        106 CTTCTTTTAAGAACAATGTTACATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGA 47

Qy       6138 TTTAAATAAACATATTTG 6155
              ||| |||||| |||||||
Db         46 TTTCAATAAAAATATTTG 29


RESULT 12
AI225639
LOCUS       AI225639              610 bp    mRNA    linear   EST 29-OCT-1998
DEFINITION  uj13g09.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:1907968 5' similar to TR:Q62766 Q62766 SSECKS. ;, mRNA
            sequence.
ACCESSION   AI225639
VERSION     AI225639.1  GI:3808692
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:976164
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1..610
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1907968"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT     172 a    163 c    185 g     89 t      1 others
ORIGIN

  Query Match             8.1%;  Score 501.8;  DB 9;  Length 610;
```

```
  Best Local Similarity   88.9%;  Pred. No. 6.2e-103;
  Matches  542;  Conservative    0;  Mismatches   68;  Indels    0;  Gaps    0;

Qy    2795 CATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGA 2854
Db       1 CACTGCAGAAGCAACTCCTGCACTCGCCCAGACTTTACCAGGGGGCAAAGATGCCCATGA 60

Qy    2855 CGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTC 2914
Db      61 CGACATAGTCACCAGTGAAGTGGATTTTACCTCAGAAGCAGTGACAGCCGCAGAAACCAC 120

Qy    2915 AGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACAT 2974
Db     121 AGAGGCGCTCCGCGCTGAAGAACTTACCGAAGCATCAGGGGCAGAAGAGACCACAGACAT 180

Qy    2975 GGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGT 3034
Db     181 GGTGTCTGCAGTTTCCCAGCTGTCCGACTCCCCGGACACCACAGAGGAAGCCACCCCAGT 240

Qy    3035 TCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCAT 3094
Db     241 TCAGGAGGTAGAGGGTGGCATGCTAGATACGGAAGAACAGGAGCGCCAGACGCAGGCCGT 300

Qy    3095 CCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGT 3154
Db     301 CCTCCAAGCCGTTGCAGACAAAGTGAAAGAGGACTCCCAGGTGCCTGCAACCCAGACTCT 360

Qy    3155 GCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGT 3214
Db     361 GCAGAGAGCAGGGCCGAAAGCACTGGAGAAGGTGGAGGAGGGTAGAGAGGACTTCGAGGT 420

Qy    3215 GCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGC 3274
Db     421 GCTGGCTACCGAGAAAGAGAAGGATGTTGTGCCGGAAGGACCCGTGCAGGAAGCTNGAAC 480

Qy    3275 TGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCC 3334
Db     481 TGAGCATCTTGCACAGGGCTTCGAGACTGGACCAGCTACCCCAGAGAGCCTTTGAGTTCC 540

Qy    3335 TGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCT 3394
Db     541 TGGAGTCACAGAGGATGTAGACCGTGCCACCACATGCCAGGTTATCAAGCACCAGCAGCT 600

Qy    3395 GATGGAACAG 3404
Db     601 GATGGAACAG 610


RESULT 13
BM228830/c
LOCUS       BM228830                 665 bp    mRNA    linear   EST 31-JAN-2002
DEFINITION  K0268G02-3 NIA Mouse Unfertilized Egg cDNA Library (Long) Mus
            musculus cDNA clone K0268G02 3', mRNA sequence.
ACCESSION   BM228830
VERSION     BM228830.1  GI:17791601
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 665)
  AUTHORS   Piao,Y., Kargul,G.J., Dudekula,D.B., Qian,Y., Luo,A., Stagg,C.A.,
            Martin,P., Aiba,K., Tanaka,T. and Ko,M.S.H.
  TITLE     Systematic Analyses of NIA Mouse Unfertilized Egg cDNA Library
            (Long)
  JOURNAL   Unpublished (2001)
COMMENT     Contact: Dawood B. Dudekula
            Laboratory of Genetics
            National Institute on Aging/National Institutes of Health
            333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
            Email: cdna@lgsun.grc.nia.nih.gov
            Plate: K0268  row: G  column: 02
            Seq primer: -21M13 Forward
            High quality sequence stop: 665
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..665
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="niaEST:K0268G02-3"
                     /db_xref="taxon:10090"
                     /clone="K0268G02"
                     /clone_lib="NIA Mouse Unfertilized Egg cDNA Library (Long
                     )"
                     /tissue_type="Unfertilized Egg"
                     /lab_host="DH10B"
                     /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2:
                     NotI; Mouse cDNA project by the Laboratory of Genetics,
                     National Institute on Aging (NIA), Intramural Research
                     Program, NIH (http://lgsun.grc.nia.nih.gov/cDNA). This is
                     a long-transcript enriched cDNA library (Ref. Genome Res.
                     11: 1553-1558 (2001). [PMID: 11544199]). Total RNAs were
                     extracted from a pool of 1488 unfertilized eggs.
                     Double-stranded cDNAs were synthesized with an Oligo(dT)
                     primer [Invitrogen:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT-3'],
                     treated with T4 DNA polymerase, and purified by
                     ethanol-precipitation. The cDNAs were ligated to
                     Lone-linker LL-Sal4, purified by phenol/chloroform, and
                     separated from free linkers by Centricon 100. Then, the
                     cDNAs were amplified by long-range high fidelity PCR using
                     Ex Taq polymerase (Takara) with a primer Sal4-S. The
                     products were purified by phenol/chloroform and Centricon
                     100. The cDNAs were digested with SalI and NotI enzymes
                     and cloned into SalI/NotI site of pSPORT1 plasmid vector.
                     The DH10B E. coli host was transformed with the ligation
                     mixture by the standard chemical method. The average
                     insert size is about 2.5 kb. The library was constructed
                     by Yulan Piao (NIA)."
BASE COUNT      237 a    126 c    120 g    182 t
ORIGIN

  Query Match             8.0%;  Score 492.2;  DB 13;  Length 665;
  Best Local Similarity   90.3%;  Pred. No. 9.4e-101;
  Matches  606;  Conservative    0;  Mismatches   53;  Indels   12;  Gaps    7;

Qy    5489 TGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATTGTATA-TTTTTTTCT 5547
Db     665 TGTATTTATATGTATGTTTTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTTCT 606

Qy    5548 AACGTTTAAGCACATGCTTTTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGA 5607
Db     605 AAGGTTTAAGCACATGCCTTTTGTATTATGCAGTATATAATGGGTGTGCAGCCATAGCGA 546

Qy    5608 CGCTTTGAAAAGCTCCAAGCCTCAACTGTAACCTGCAGCAAACAG----ATAACATTCCT 5663
Db     545 AGCTTTGAGAAGCCCCAAGCCTCAACTGTAACCTGCAGCAAACAGAGAAATAACATTCCT 486

Qy    5664 GGCAAGAAGAGACAAGTCTTTTTTAAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGT 5723
Db     485 GGCAGGAAGATACAAGTC-TTTTTAAAGTTTACTGATGCTTAACTCTGTGGGCTCGTAGT 427

Qy    5724 CCTCTGAAAGTGGTTGTTTTCCTATGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGA 5783
Db     426 CCTCTGAAAGTGGTTGTTTTCCTATGCACAGTGAGCTCAGAAATAAAAACTCCATTTTGA 367

Qy    5784 AACATCCAGGATGTCCCAATATTACCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCC 5843
Db     366 GACATTCAGAATGTCCCAACATTACGACAACATTTTTT---TTTTTTTCTAATCCAGTCC 310

Qy    5844 AGGTTGGAAAGAAGTCTCCTCTGTGTCAGATTAAGCCCTGTCTCTTAATGATATGGACAA 5903
Db     309 AGGTTGGAAAGAAGTCTCCTTAGTGTCAGATTAAGCCCCATCTCTTAACAATATGGACAG 250

Qy    5904 ATGAGTGTGCCTAAGGCCATGAGATGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTT 5963
```

```
Db      249 ATGAGTGTGCC-ATGGCCATGAGCTATTTCCTAACGCAGAAGGAATTTGTTGT-TACTTT 192

Qy     5964 TTTGATTGTACTCTTCTATGCTGGACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTC 6023
            || |||||||||||||| ||||||||||||||||||||||| |||||||| |||| ||
Db      191 TTGGATTGTACTCTTCTAGGCTGGACCGAATTCATATGCAGATTGAAGTGAGACCTGCTC 132

Qy     6024 TTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTC- 6082
            ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db      131 TTTACAGATGGTATTTTGATAGATACTGGAGTTTGTCTGTATTATATCTGTGCCCCTTCT 72

Qy     6083 TTTAAGAACAATGTTGCATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTTAA 6142
            |||||||||||||| |||||||||||||||||||||||||||||||||||||||||| |
Db       71 TTTAAGAACAATGTTACATTATGTTCCTTTGGATAAATTGTGATTTGACAACTGATTTCA 12

Qy     6143 ATAAACATATT 6153
            ||||| || ||
Db       11 ATAAAAATTTT 1
```

```
RESULT 14
AI103489/c
LOCUS       AI103489                 476 bp    mRNA    linear   EST 08-JAN-1999
DEFINITION  EST212778 Normalized rat embryo, Bento Soares Rattus sp. cDNA clone
            REMCC63 3' end, mRNA sequence.
ACCESSION   AI103489
VERSION     AI103489.1  GI:4134176
KEYWORDS    EST.
SOURCE      Rattus sp.
  ORGANISM  Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Lee,N.H., Glodek,A., Chandra,I., Mason,T.M., Quackenbush,J.,
            Kerlavage,A.R. and Adams,M.D.
  TITLE     Rat Genome Project: Generation of a Rat EST (REST) Catalog & Rat
            Gene Index
  JOURNAL   Unpublished (1998)
COMMENT     On Oct 6, 1998 this sequence version replaced gi:3708028.
            Contact: Lee, NH
            The Institute for Genomic Research
            9712, Medical Center Drive, Rockville, MD 20850, USA
            Tel: (301)-838-3529
            Fax: (301)-838-0208
            Email: nhlee@tigr.org
            Seq primer: M13-21.
FEATURES             Location/Qualifiers
     source          1..476
                     /organism="Rattus sp."
                     /db_xref="taxon:10118"
                     /clone="REMCC63"
                     /clone_lib="Normalized rat embryo, Bento Soares"
                     /dev_stage="embryo 8, 12, 18 dpc"
                     /note="Vector: pT7T3Pac; Site_1: EcoRI; Site_2: NotI"
BASE COUNT      170 a     94 c     90 g    122 t
ORIGIN

  Query Match             7.6%;  Score 467.8;  DB 9;  Length 476;
  Best Local Similarity   99.6%;  Pred. No. 3.1e-95;
  Matches  469;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy     5688 AAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTA 5747
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      476 AAAGTTTACTGATGCTTAGATCTGTGGGCTTCTAGTCCTCTGAAAGTGGTTGTTTTCCTA 417

Qy     5748 TGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTA 5807
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      416 TGCACAGCGAGCTCAGAAATAAAAACCCCATTTTGAAACATCCAGGATGTCCCAATATTA 357

Qy     5808 CCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGT 5867
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      356 CCATGATTTTTTCCCCCCTTTTTGCTAATCCAGTCCAGGTTGGAAAGAAGTCTCCTCTGT 297

Qy     5868 GTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGA 5927
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      296 GTCAGATTAAGCCCTGTCTCTTAATGATATGGACAAATGAGTGTGCCTAAGGCCATGAGA 237

Qy     5928 TGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTATGCTGG 5987
            |||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db      236 TGTTTCCTAATGCAGAAGGAATCTGTTGTACGTTTTTTTGATTGTACTCTTCTAGGCTGG 177

Qy     5988 ACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGAT 6047
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      176 ACCGAATTCATATGCAGATCGAAGTGAGTCCTGTTCTTTACAGATGGTATTTTGATAGAT 117

Qy     6048 ACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTT 6107
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      116 ACTGGAGTTTGTCTGTGTTATATCTGTGCCCCTTCTTTAAGAACAATGTTGCATTATGTT 57

Qy     6108 CCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGACT 6158
            ||||||||||||||||||||||||||||||||||||||||||||||||  ||
Db       56 CCTTTGGATAAATTGTGATTTGACAACTGATTTAAATAAACATATTTGCCT 6
```

```
RESULT 15
BG670384
LOCUS       BG670384                 573 bp    mRNA    linear   EST 30-APR-2001
DEFINITION  DRNBAE11 Rat DRG Library Rattus norvegicus cDNA clone DRNBAE11 5',
            mRNA sequence.
ACCESSION   BG670384
VERSION     BG670384.1  GI:13892483
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 573)
  AUTHORS   Xiao,H.S., Huang,Q.H., Zhang,F.X., Bao,L., Lu,Y.J., Guo,C., Yang,L.
            , Huang,W.J., Fu,G., Xu,S.H., Cheng,X.P., Yan,Q., Zhu,Z.D., Zhang
            ,X., Chen,Z., Han,Z.G. and Zhang,X.
  TITLE     Identification of gene expression profile of dorsal root ganglion
            in the rat peripheral axotomy model of neuropathic pain
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (12), 8360-8366 (2002)
  MEDLINE   22056133
COMMENT     Contact: Zhang Xu
            Laboratory of Sensory System
            Institute of Neuroscience
            320 Yue Yang Road, Shanghai 200031, P.R.China
            Tel: 86-21-64748700-121
            Fax: 86-21-64713446
            Email: xu.zhang@ion.ac.cn
            This clone is also available at Chinese National Human Genome
            Center at Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park,
            Pudong New Area, P.R.China. Please contact with Zhang Xu
            (xu.zhang@ion.ac.cn) or Han Zeguang (hanzg@chgc.sh.cn)
            PCR PRimers
            FORWARD: T3
            BACKWARD: T7
            Seq primer: T3
            POLYA=No.
FEATURES             Location/Qualifiers
     source          1..573
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="DRNBAE11"
                     /clone_lib="Rat DRG Library"
                     /sex="male"
                     /tissue_type="dorsal root ganglion"
                     /dev_stage="adult"
                     /note="Total RNA was isolated from hypothalamus and
                     transcripted into cDNA,  which was then used as templat in
                     PCR.The PCR products were separated  on sequencing gel.
                     The differential bands were cut, reamplified, cloned into
                     pMD18-T vector and confirmed by Northern blot."
```

```
BASE COUNT      151 a    131 c    115 g    170 t      6 others
ORIGIN

  Query Match             7.5%;  Score 460.2;  DB 12;  Length 573;
  Best Local Similarity  88.0%;  Pred. No. 1.7e-93;
  Matches  557;  Conservative    0;  Mismatches   14;  Indels   62;  Gaps    3;

Qy       5088 GAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTC 5147
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 GAATCCTAAGATGTTAGTTGCTCATTGTACATCTGTAAGACCAGAATGTGAAAACAAGTC 60

Qy       5148 ACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCA 5207
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACAGAACAAGATGCTGCTGTTGGGACCTTGAGACCAAGATTTCAGAGCCCATGAGATCCA 120

Qy       5208 GAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATTCTGAGGCTTC 5267
              ||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        121 GAGAGCAGGGCCGTCCAATGATTTCCACCCAGTAGAGCACCCCGACAATCCTGAGGCTTC 180

Qy       5268 ATCGGGAGCTAGAGCCAGCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTT 5327
              ||||||||||||||||||
Db        181 ATCGGGAGCTAGAGCCAG------------------------------------------ 198

Qy       5328 GATGCCGTCCGTGTATTTCTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTT 5387
                                ||||||||||||||||||||||||||||||||||||||||||
Db        199 ------------------CTAACATTTCCTCGTTTCAAGACTGCCTTTGATTTGCCCCTT 240

Qy       5388 GATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAAT-TCTG 5446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        241 GATGCCGTCCGTGTATTTCGGATTTAAGGTCCTGCGTTCTCAACCTGGAACCAATGTCTG 300

Qy       5447 CCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTT 5506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 CCATACCTAGTTCCACTTCTCAAACTGGAGCATCCTCCTTTATGTATTTATATGTATGTT 360

Qy       5507 TTATGTAGTCCTCCTCCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTT 5566
               ||||||| |||||| ||||||||||||||||||||||||||||||||||||||||||||
Db        361 NTATGTAGNCCTCCTNCTGTACCTATTGTATATTTTTTTCTAACGTTTAAGCACATGCTT 420

Qy       5567 TTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTCCAAG 5626
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db        421 TTTGTATTATGCAATATATAACGGGTGTGCAGCCATAGCGACGCTTTGAAAAGCTNCAAG 480

Qy       5627 CCTCAACTGTAACCTGCAGCAAACAGATAACA-TTCCTGGCAAGAAGAGACAAGTCTTTT 5685
              ||||||||||||||||||||||||||||| ||| ||||||||||     |   |||||
Db        481 CCTCAACTGTAACCTGCAGCAAACAGATAACATTTCTTGGCAAGAAGAAGACAANCTTTT 540

Qy       5686 TTAAAGTTTACTGATGCTTAGATCTGTGGGCTT 5718
              ||||||  |||||||||||||||||||||||||
Db        541 TTAAAGGNTACTGATGCTTAGATCTGTGGGCTT 573

Search completed: December 12, 2002, 16:19:41
Job time : 7377 secs
```

```
OM protein - protein search, using sw model

Run on:          December 12, 2002, 16:39:22 ; Search time 123 Seconds
                                               (without alignments)
                                               1729.008 Million cell updates/sec

Title:           US-09-902-432-4
Perfect score:   8073
Sequence:        1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters:      908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       A_Geneseq_101002:*
                 1:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
                 2:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
                 3:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
                 4:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
                 5:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
                 6:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
                 7:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
                 8:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
                 9:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
                 10:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
                 11:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
                 12:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
                 13:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
                 14:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
                 15:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
                 16:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
                 17:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
                 18:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
                 19:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
                 20:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
                 21:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
                 22:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
                 23:  /SIDS2/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8049	99.7	1596	18	AAW31347	Rat tumour suppres
2	5387.5	66.7	1346	18	AAW31346	Rat tumour suppres
3	4334	53.7	1783	23	ABB97448	Novel human protei
4	4294	53.2	1781	23	AAO17365	Human gravin. Hom
5	4274.5	52.9	1780	19	AAW53863	Human gravin polyp
6	4274.5	52.9	1780	21	AAB15380	Human gravin prote
7	4262.5	52.8	1795	22	ABG21018	Novel human diagno
8	952.5	11.8	292	19	AAW53862	Gravin polypeptide
9	952.5	11.8	292	21	AAB15379	Recombinant human
10	656	8.1	651	19	AAW53875	Gravin polypeptide
11	656	8.1	651	21	AAB18410	Human gravin PKA r
12	642	8.0	2768	22	ABB68397	Drosophila melanog
13	639.5	7.9	453	19	AAW53867	Gravin polypeptide
14	639.5	7.9	453	21	AAB18407	Human gravin PKA r
15	557.5	6.9	6815	22	ABB66811	Drosophila melanog
16	521.5	6.5	396	19	AAW53868	Gravin polypeptide
17	521.5	6.5	396	21	AAB18408	Human gravin PKA r
18	492	6.1	2519	22	ABG16636	Novel human diagno
19	485	6.0	3263	22	ABB67210	Drosophila melanog
20	462	5.7	1786	18	AAW24790	P. falciparum live
21	462	5.7	1787	23	AAU96699	Plasmodium falcipo
22	451	5.6	1468	22	ABB62991	Drosophila melanog
23	440.5	5.5	1812	22	ABB58022	Drosophila melanog
24	431	5.3	3257	22	ABB67502	Drosophila melanog
25	430	5.3	3111	22	ABB60327	Drosophila melanog
26	429	5.3	2665	22	ABB28314	Human peptide #965
27	429	5.3	2665	22	ABB33490	Peptide #996 encod
28	429	5.3	2665	22	ABB18950	Protein #949 encod
29	429	5.3	2665	22	AAM54270	Human brain expres
30	429	5.3	2665	22	AAM66665	Human bone marrow
31	429	5.3	2665	22	AAM14533	Peptide #967 encod
32	429	5.3	2665	22	AAM26950	Peptide #987 encod
33	429	5.3	2665	22	AAM02259	Peptide #941 encod
34	429	5.3	2665	23	ABG36319	Human peptide enco
35	429	5.3	2748	22	ABB58843	Drosophila melanog
36	429	5.3	3266	21	AAB42491	Human ORFX ORF2255
37	421	5.2	1714	22	ABG02138	Novel human diagno
38	421	5.2	1714	22	ABG06417	Novel human diagno
39	420	5.2	2803	23	ABB08161	Human cytoskeleton
40	414.5	5.1	2622	22	ABG06418	Novel human diagno
41	411	5.1	2618	22	ABG02135	Novel human diagno
42	409.5	5.1	2515	22	ABB12281	Human secreted pro
43	409.5	5.1	2515	22	AAM80268	Human protein SEQ
44	409.5	5.1	2515	22	AAM80269	Human protein SEQ
45	409	5.1	1793	22	ABB59613	Drosophila melanog

ALIGNMENTS

```
RESULT 1
AAW31347
ID   AAW31347 standard; Protein; 1596 AA.
XX
AC   AAW31347;
XX
DT   20-JUL-1998  (first entry)
XX
DE   Rat tumour suppressor protein SSeCKS.
XX
KW   SSeCKS; tumour suppressor gene; rat; protein kinase C; mitosis;
KW   cancer; malignancy; cell proliferation; Alzheimer's disease;
KW   therapy.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..10
FT                   /note= "myristylation site"
FT   Region          410..418
FT                   /note= "glycine-rich region"
FT   Peptide         517..520
FT                   /note= "nuclear localisation signal"
FT   Region          524..527
FT                   /note= "glycine-rich region"
FT   Peptide         677..680
FT                   /note= "nuclear localisation signal"
FT   Peptide         757..760
FT                   /note= "nuclear localisation signal"
FT   Region          1399..1421
FT                   /note= "zinc finger"
FT   Modified-site   279..307
FT                   /note= "proposed protein kinase C phosphorylation
```

```
FT                        site"
FT   Modified-site   504..526
FT                   /note= "proposed protein kinase C phosphorylation
FT                        site"
FT   Modified-site   592..614
FT                   /note= "proposed protein kinase C phosphorylation
FT                        site"
FT   Modified-site   741..766
FT                   /note= "proposed protein kinase C phosphorylation
FT                        site"
XX
PN   WO9740059-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;   97WO-US06830.
XX
PR   18-JUN-1996;   96US-0665401.
PR   19-APR-1996;   96US-0635121.
XX
PA   (GELM/) GELMAN I.
PA   (JAKE/) JAKEN S.
XX
PI   Gelman I,  Jaken S;
XX
DR   WPI; 1997-535770/49.
DR   N-PSDB; AAV02302.
XX
PT   Tumour suppressor gene SSeCKS - used as a mitotic regulator, and
PT   inhibitor of malignant phenotype
XX
PS   Claim 4; Fig 11A-L; 162pp; English.
XX
CC   This polypeptide comprises the novel tumour suppressor protein
CC   SSeCKs that is a substrate of protein kinase C and which acts as a
CC   negative regulatory of mitosis and as an inhibitor of the
CC   transformed phenotype.  Its amino acid sequence was deduced from an
CC   isolated nucleic acid molecule (see AAV02302).  The SSeCKS nucleic
CC   acid, as well as homologous and hybridising nucleic acids are
CC   claimed, as are isolated proteins encoded by such nucleic acids,
CC   vectors comprising the nucleic acids, host cells, and methods of
CC   inhibiting the expression of a transformed phenotype in a host cell
CC   by introducing the nucleic acids.  Introduction of a SSeCKS nucleic
CC   acid or gene product into a host cell inhibits mitosis of the host
CC   cell, allowing the treatment of diseases associated with disorders
CC   of proliferation and/or with the expression of a malignant
CC   phenotype. SSeCKS can also be used to treat or identify disorders
CC   of cytoskeletal structure and cellular architecture (such as
CC   Alzheimer's disease), and may be a marker for aberrancies in
CC   fertility and/or nervous system development.
XX
SQ   Sequence   1596 AA;

  Query Match             99.7%;  Score 8049;  DB 18;  Length 1596;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 1593;  Conservative    0;  Mismatches    3;  Indels    0;  Gaps    0;

Qy          1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60

Qy         61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120

Qy        121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy        181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240

Qy        241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
              ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db        241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSSKKFFTHGWAGWRKKTS 300

Qy        301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360

Qy        361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420

Qy        421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480

Qy        481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTE 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTE 540

Qy        541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600

Qy        601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Qy        661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720

Qy        721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
              |||||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db        721 EQDQAQGSSSPEPAGSLSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780

Qy        781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840

Qy        841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Qy        901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Qy        961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Qy       1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080

Qy       1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140

Qy       1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200

Qy       1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260

Qy       1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320

Qy       1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380
```

```
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380

Qy     1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
            ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     1381 AVEKVIETVVISETGESPECVGDHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440

Qy     1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500

Qy     1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560

Qy     1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
            ||||||||||||||||||||||||||||||||||||
Db     1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596


RESULT 2
AAW31346
ID   AAW31346 standard; Protein; 1346 AA.
XX
AC   AAW31346;
XX
DT   20-JUL-1998  (first entry)
XX
DE   Rat tumour suppressor protein SSeCKS (active truncated form).
XX
KW   SSeCKS; tumour suppressor gene; rat; protein kinase C; mitosis;
KW   cancer; malignancy; cell proliferation; Alzheimer's disease;
KW   therapy.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   Misc-difference 72
FT                   /note= "encoded by TCT"
FT   Misc-difference 496
FT                   /note= "encoded by GTC"
FT   Misc-difference 499
FT                   /note= "encoded by GTC"
FT   Misc-difference 757
FT                   /note= "encoded by CCC"
FT   Misc-difference 785
FT                   /note= "encoded by AGG"
FT   Region          24..32
FT                   /note= "glycine-rich region"
FT   Peptide         131..134
FT                   /note= "nuclear localisation signal"
FT   Region          138..141
FT                   /note= "glycine-rich region"
FT   Peptide         202..205
FT                   /note= "nuclear localisation signal"
FT   Peptide         289..292
FT                   /note= "nuclear localisation signal"
FT   Region          306..315
FT                   /note= "glycine-rich region"
FT   Peptide         368..371
FT                   /note= "nuclear localisation signal"
FT   Region          1013..1035
FT                   /note= "zinc finger"
XX
PN   WO9740059-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;   97WO-US06830.
XX
PR   18-JUN-1996;   96US-0665401.
PR   19-APR-1996;   96US-0635121.
XX
PA   (GELM/) GELMAN I.
PA   (JAKE/) JAKEN S.
XX
PI   Gelman I,  Jaken S;
XX
DR   WPI; 1997-535770/49.
DR   N-PSDB; AAV02301.
XX
PT   Tumour suppressor gene SSeCKS - used as a mitotic regulator, and
PT   inhibitor of malignant phenotype
XX
PS   Example 6; Fig 3A-G; 162pp; English.
XX
CC   This polypeptide comprises an active truncated form of the novel
CC   tumour suppressor protein SSeCKs that is a substrate of protein
CC   kinase C and which acts as a negative regulatory of mitosis and as
CC   an inhibitor of the transformed phenotype.  Its amino acid sequence
CC   was deduced from a partial cDNA clone (see AAV02301) obtained from
CC   NIH3T3 cells.  The full-length SSeCKS sequence (see AAW31347) is also
CC   provided.  Full-length SSeCKS nucleic acid (see AAV02301), as well as
CC   homologous and hybridising nucleic acids are claimed, as are
CC   isolated proteins encoded by such nucleic acids, vectors comprising
CC   the nucleic acids, host cells, and methods of inhibiting the
CC   expression of a transformed phenotype in a host cell by introducing
CC   the nucleic acids.  Introduction of a SSeCKS nucleic acid or gene
CC   product into a host cell inhibits mitosis of the host cell,
CC   allowing the treatment of diseases associated with disorders of
CC   proliferation and/or with the expression of a malignant phenotype.
CC   SSeCKS can also be used to treat or identify disorders of
CC   cytoskeletal structure and cellular architecture (such as
CC   Alzheimer's disease), and may be a marker for aberrancies in
CC   fertility and/or nervous system development.
XX
SQ   Sequence   1346 AA;

  Query Match             66.7%;  Score 5387.5;  DB 18;  Length 1346;
  Best Local Similarity   91.1%;  Pred. No. 1.8e-279;
  Matches 1092;  Conservative   25;  Mismatches   67;  Indels   15;  Gaps    7;

Qy      387 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 446
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 60

Qy      447 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 506
            |||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 EELMKSREMCVEGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 120

Qy      507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 180

Qy      567 EKGPLEAPQDGEAEEGTTSDGEKKRE----GITPWASFKKMVTPKKRVRRPSESDKEEEL 622
            ||||||||: |:  :       |||||:      ||       :|    :: |||||||||||
Db      181 EKGPLEAPRMGKLRKELLR-GEKKRKDHSLGI-----LQKDGDTQETVRRPSESDKEEEL 234

Qy      623 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 682
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      235 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 294

Qy      683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 742
            ||||    ||||||     ::       : : : ||||||||||||||||||||||||||||
Db      295 SSSDIR-GPRTLGGGQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 353

Qy      743 VSTWESFKRLVTPRKKSKSKLEEK-AEDSSVEQLSTEIEPSREESWVS--IKKFIPGRRK 799
            |||||||||||||||||||||| |  : |      :  :| |   :    ::   |    :
Db      354 VSTWESFKRLVTPRKKSKSKLEEKEAGRTLVVGAGCPLRSNRVEKNLGFPLRNSSPDGGR 413

Qy      800 KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVY 859
            |     |:||||||||||||||:|:||||||||||:||||||||||| |||    |  |
Db      414 KGQMGRQEQATVEDSGPVEINEDEPDVPAVVPLSEYDAVEREKMEAQGNAELPSCWGCV- 472

Qy      860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919
            |||||||||||||||||||||| || |||||||||||||||||||||||||||||||||
```

```
Db     473 VSEELSKTLVHTVSVAVIDGTRAKTSKEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 532

Qy     920 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 979
               |||||||||||||||||:||||||||||||||||||||||||||||||||||||||||||
Db     533 DIIAEETPVLTQTLPEGQDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 592

Qy     980 DMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039
               ||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||
Db     593 DMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 652

Qy    1040 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 1099
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     653 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 712

Qy    1100 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 1159
               |||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db     713 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTRLADSDTADGTQQDET 772

Qy    1160 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELT 1219
               |||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||| 
Db     773 IDSQDSKATAAVEQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELA 832

Qy    1220 AAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
               ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     833 AAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 892

Qy    1280 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPS 1339
               ||||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db     893 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVLTLDMPS 952

Qy    1340 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     953 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1012

Qy    1400 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1459
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1013 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1072

Qy    1460 ASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1519
               ||||||||||||||||||||||||||||:|||||||||||||||||||||||||||||||
Db    1073 ASQRERSEEEDKPDAGPDADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1132

Qy    1520 RTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAW 1578
               ||||||||||||||||||| ||||||||||||||| ||||||||||||||||||||||| 
Db    1133 RTETAPETHAYDSQTQVPAMRLDSREPNRCWTKMKVAKMKHPVPQPREDLQVLTVLEAW 1191
```

RESULT 3
ABB97448
ID   ABB97448 standard; Protein; 1783 AA.
XX
AC   ABB97448;
XX
DT   27-JUN-2002 (first entry)
XX
DE   Novel human protein SEQ ID NO: 716.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
XX
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
PI   Xue AJ, Yang Y, Wehrman T, Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   N-PSDB; ABN32634.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Claim 20; SEQ ID NO 716; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a protein of the invention.
XX
SQ   Sequence   1783 AA;

```
  Query Match             53.7%;  Score 4334;  DB 23;  Length 1783;
  Best Local Similarity   58.2%;  Pred. No. 5.6e-223;
  Matches  962;  Conservative  188;  Mismatches  401;  Indels  102;  Gaps   32;

Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
             ||||||||||||| | || ||:|   || ||:|||:     ||| |  |||||| ||||
Db       1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy      57 QLSSVNGVAEQGDVHVQEENQEGQ-------------EEEVVDEDVGQRESEDVREKDR 102
             |||::||||| :: :|| :  ||              ||||:  :||||:||||  |:|
Db      61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy     103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
              :|||  |  |||  ||||||  ||||||:||:|:|:  ||  |||||:|||||||||
Db     121 DKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 180

Qy     163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
             |||||||| || ||||||||||||||||   ||||||:||:      ||:|||||| |||
Db     181 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 235

Qy     223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
             ||| | |||:|||   ||    |||  ||   || |:|||||||||:|| |||:||| |
Db     236 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 294

Qy     283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
              :||||| ||||||||||:| ||::| :||:|||| ||||  || |  | ||:
Db     295 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 354

Qy     338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
                 |||||   :  ||||:||||||| |:||   :   ||| ||||||||||:|  | |
Db     355 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 414

Qy     393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
             ||||||||||:  |||        |  ||  |:|| |:||  |  |||||||:|:|:|:|
Db     415 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDAEPQEAEPAKELVKL 466

Qy     453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
             :| |||  ||  ||||:||  |  |||:||||||||||||||:|||||||||||:|:|||
Db     467 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 526

Qy     513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
             ||||||||||| ||||| ||:  :  :|| :|||||||||||||||||| |||||||  |
Db     527 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 585

Qy     573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
              ||||||||| ||||||||||:||||||||||||||||||||||||||:||:||||||||
Db     586 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 645
```

```
Qy    633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692

Db    646 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPK 705

Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752

Db    706 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 765

Qy    753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808

Db    766 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 825

Qy    809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865

Db    826 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 885

Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924

Db    886 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 945

Qy    925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984

Db    946 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1005

Qy    985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038

Db   1006 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1065

Qy   1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097

Db   1066 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1119

Qy   1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147

Db   1120 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1179

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207

Db   1180 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1238

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259

Db   1239 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1296

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314

Db   1297 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1356

Qy   1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372

Db   1357 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1416

Qy   1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430

Db   1417 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1476

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490

Db   1477 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEEDL 1534

Qy   1491 KAEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRC 1549

Db   1535 EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQE 1594

Qy   1550 WTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577

Db   1595 TEKEGEDPQASAQDETPITSAKEESESTAVGQA 1627
```

RESULT 4
AAO17365
ID   AAO17365 standard; protein; 1781 AA.
XX
AC   AAO17365;
XX
DT   19-JUL-2002 (first entry)
XX
DE   Human gravin.
XX
KW   Human; endometriosis; DNA chip; fibronectin; p27; reticulocalbin;
KW   aldehyde dehydrogenase 6; gravin; phospholipase C epsilon; elastin;
KW   insulin-like growth factor binding protein-2; alpha-2 type IV collagen;
KW   transmembrane receptor PTK7; collagen type XVIII alpha 1;
KW   platelet derived growth factor receptor alpha; laminin M chain;
KW   subtilisin like protein PACE4; nidogen.
XX
OS   Homo sapiens.
XX
PN   EP1191107-A2.
XX
PD   27-MAR-2002.
XX
PF   21-AUG-2001; 2001EP-0250300.
XX
PR   25-SEP-2000; 2000DE-1048633.
XX
PA   (SCHD) SCHERING AG.
XX
PI   Hess-stumpp H, Haendler B, Kraetzschmar J, Kreft B, Winterhager E;
PI   Regidor P, Scotti S;
XX
DR   WPI; 2002-317413/36.
XX
PT   In vitro diagnosis and monitoring of endometriosis, comprises
PT   detecting reduced expression of specific gene products, e.g. from the
PT   fibronectin gene -
XX
PS   Claim 1; Page 18-19; 21pp; German.
XX
CC   The present invention relates to a method for the in vitro diagnosis of
CC   endometriosis by determining the amount of gene product from at least one
CC   specific gene in a patient sample and comparing this with the amount of
CC   gene product in a control sample. A reduced level is indicative of
CC   endometriosis. The gene products may be fibronectin, p27, reticulocalbin,
CC   aldehyde dehydrogenase 6, gravin, phospholipase C epsilon, elastin,
CC   insulin-like growth factor binding protein-2, alpha-2 type IV collagen,
CC   transmembrane receptor PTK7, collagen type XVIII alpha 1,
CC   platelet derived growth factor receptor alpha, laminin M chain,
CC   subtilisin like protein PACE4 or nidogen. The method is useful for
CC   initial diagnosis of endometriosis, and also for monitoring progress and
CC   treatment of the disease. The present sequence is human gravin.
XX
SQ   Sequence   1781 AA;

```
  Query Match             53.2%;  Score 4294;  DB 23;  Length 1781;
  Best Local Similarity   57.9%;  Pred. No. 7.8e-221;
  Matches  959;  Conservative  187;  Mismatches  401;  Indels  108;  Gaps   35;

Qy      1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56

Db      1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy     57 QLSSVNGVAEQGDVHVQEENQEGQ-------------EEEVVDEDVGQRESEDVREKDR 102

Db     61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy    103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162

Db    121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy    163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222

Db    180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy    223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
```

```
Db    235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy    283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db    294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy    338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db    354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy    393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db    414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy    453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db    466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy    513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db    526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy    573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEELEKVKSATLSS 632
Db    585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy    633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db    645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPK 704

Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db    705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy    753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSRREESWVSIKKFIPGRRKKRADGKQEQ 808
Db    765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy    809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db    825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db    885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy    925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db    945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy    985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db   1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy   1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db   1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy   1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db   1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db   1179 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1237

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db   1238 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1295

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db   1296 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1355

Qy   1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db   1356 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1415

Qy   1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db   1416 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1475

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db   1476 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1532

Qy   1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547
Db   1533 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1590

Qy   1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db   1591 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1625
```

```
RESULT 5
AAW53863
ID   AAW53863 standard; peptide; 1780 AA.
XX
AC   AAW53863;
XX
DT   13-JUL-1998  (first entry)
XX
DE   Human gravin polypeptide.
XX
KW   Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW   cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW   Myasthenia gravis; nicotinic acetylcholine receptor.
XX
OS   Homo sapiens.
XX
PN   US5741890-A.
XX
PD   21-APR-1998.
XX
PF   19-DEC-1996;   96US-0769309.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Nauert JB,  Scott JD;
XX
DR   WPI; 1998-260552/23.
DR   N-PSDB; AAV23545.
XX
PT   New polypeptide fragments of protein kinase binding protein gravin -
PT   are useful for the study of modulation of action between gravin and
PT   protein kinase(s)
XX
PS   Example 1; Column 19-32; 32pp; English.
XX
CC   This sequence corresponds to the human gravin polypeptide, and represents
CC   a polypeptide of the invention. The polypeptides are fragments capable of
CC   binding to type II regulatory subunit of cAMP-dependent protein kinase
CC   (PKA). Gravin is a kinase anchoring protein that binds to type II
CC   regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC   antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC   develops antibodies against their own nicotinic acetylcholine receptors.
CC   The polypeptides are useful for providing analogues of gravin in the
CC   study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC   interactions between gravin and kinase. The peptides are involved in the
CC   modulation of gravin-kinase interactions.
XX
SQ   Sequence   1780 AA;
```

```
  Query Match             52.9%;  Score 4274.5;  DB 19;  Length 1780;
  Best Local Similarity   57.8%;  Pred. No. 8.5e-220;
  Matches  957;  Conservative  187;  Mismatches  402;  Indels  109;  Gaps   36;

Qy          1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db          1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy         57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102
Db         61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy        103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db        121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy        163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db        180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy        223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db        235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy        283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db        294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy        338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db        354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy        393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db        414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy        453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db        466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy        513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db        526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy        573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db        585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy        633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db        645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPK 704

Qy        693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db        705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy        753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
Db        765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy        809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db        825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy        866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db        885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy        925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db        945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy        985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db       1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy       1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db       1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy       1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db       1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy       1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db       1179 FDAPGTTQKDEIVEIHEENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1236

Qy       1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db       1237 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1294

Qy       1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db       1295 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1354

Qy       1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db       1355 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1414

Qy       1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db       1415 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1474

Qy       1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db       1475 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1531

Qy       1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547
Db       1532 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1589

Qy       1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db       1590 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1624
```

```
RESULT 6
AAB15380
ID   AAB15380 standard; Protein; 1780 AA.
XX
AC   AAB15380;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Human gravin protein sequence.
XX
KW   Human; gravin; PKA RII binding site; myasthenia gravis;
KW   kinase anchoring protein; cAMP dependent protein kinase.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    265..556
FT                   /note= "PKC binding site"
FT   Binding-site    1526..1582
FT                   /note= "PKA RII binding site"
FT   Region          1537..1563
FT                   /note= "PKA anchoring site"
XX
PN   US6090929-A.
XX
PD   18-JUL-2000.
XX
PF   19-DEC-1997;   97US-0994570.
XX
```

```
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Scott JD,  Nauert JB;
XX
DR   WPI; 2000-523763/47.
DR   N-PSDB; AAA74903.
XX
PT   Novel polynucleotides useful for detecting gravin in patients suffering
PT   from Myasthenia gravis encodes cAMP-dependent protein kinase-binding
PT   polypeptide and protein kinase C-binding polypeptide of gravin  -
XX
PS   Claim 1; Column 35-45; 34pp; English.
XX
CC   The present sequence is the protein sequence of human gravin. Gravin is
CC   an A-kinase anchoring protein (AKAP) which is involved in the
CC   localisation of cAMP dependent protein kinase A (PKA) via interactions
CC   between the RII binding region and the PKA regulatory subunit RII. Gravin
CC   is also an antigen found in myasthenia gravis sufferers, and it is
CC   thought that antibodies to it may be useful in modulating the binding of
CC   PKA, and thus aid in the treatment of the disease. The gravin coding
CC   sequence was isolated by first screening a human umbilical vein
CC   endothelial cell cDNA library with serum from a myasthenia gravis
CC   patient, and then searching a human heart cDNA library for sequences
CC   resembling the isolated sequence. This was done because the first
CC   sequence obtained was shown to be shorter than the full length cDNA.
XX
SQ   Sequence   1780 AA;

  Query Match             52.9%;  Score 4274.5;  DB 21;  Length 1780;
  Best Local Similarity   57.8%;  Pred. No. 8.5e-220;
  Matches  957;  Conservative  187;  Mismatches  402;  Indels  109;  Gaps   36;

Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db       1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy      57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102
Db      61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy     103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db     121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy     163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db     180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy     223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db     235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy     283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db     294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy     338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db     354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy     393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db     414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy     453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db     466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy     513 KLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db     526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy     573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db     585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy     633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db     645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPK 704

Qy     693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db     705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy     753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
Db     765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy     809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db     825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy     866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db     885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy     925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db     945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy     985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db    1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy    1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db    1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy    1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db    1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy    1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db    1179 FDAPGTTQKDEIVEIHEENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1236

Qy    1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db    1237 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1294

Qy    1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db    1295 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1354

Qy    1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db    1355 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1414

Qy    1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db    1415 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1474

Qy    1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db    1475 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1531

Qy    1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547
Db    1532 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1589

Qy    1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db    1590 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1624
```

```
RESULT 7
ABG21018
ID   ABG21018 standard; Protein; 1795 AA.
XX
AC   ABG21018;
XX
DT   18-FEB-2002  (first entry)
XX
DE   Novel human diagnostic protein #21009.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   N-PSDB; AAS85205.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 20; SEQ ID No 51377; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. ABG00010-ABG30377 represent novel human
CC   diagnostic amino acid sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO -
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   1795 AA;

  Query Match            52.8%;  Score 4262.5;  DB 22;  Length 1795;
  Best Local Similarity  57.5%;  Pred. No. 3.8e-219;
  Matches  951;  Conservative  190;  Mismatches  410;  Indels  103;  Gaps   33;

Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db       9 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 68

Qy      57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102
Db      69 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVRQVDHIDSEDVSDRDS 128

Qy     103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db     129 DTEMTTKSAVVHDITDDGQEETPDIIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 188

Qy     163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db     189 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 243

Qy     223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db     244 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 302

Qy     283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db     303 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 362

Qy     338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db     363 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 422

Qy     393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db     423 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 474

Qy     453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db     475 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 534

Qy     513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db     535 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 593

Qy     573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db     594 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 653

Qy     633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db     654 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPK 713

Qy     693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db     714 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 773

Qy     753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
Db     774 VTPGKKSKSKLEEKSEDSLAGSGVEHSTPDTEPGKEESWVSIKKFIPGLRKKRPDGKQEQ 833

Qy     809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db     834 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELS 893

Qy     866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db     894 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 953

Qy     925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db     954 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSA 1013

Qy     985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db    1014 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1073

Qy    1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db    1074 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1127

Qy    1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db    1128 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1187
```

```
Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db   1188 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1246

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db   1247 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1304

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETK-PEQVSE 1313
Db   1305 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREEYQXQSPTHVNE 1364

Qy   1314 EG-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSL 1371
Db   1365 EKLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSS 1424

Qy   1372 DTTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPL 1429
Db   1425 EASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPT 1484

Qy   1430 GPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEDKPDAGPDADGKESTAIEKV 1489
Db   1485 GPDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEED 1542

Qy   1490 LKAEPEILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNR 1548
Db   1543 LEPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQ 1602

Qy   1549 CWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db   1603 ETEKEGEDPQASAQDETPITSAKEESESTAVGQA 1636
```

```
RESULT 8
AAW53862
ID   AAW53862 standard; peptide; 292 AA.
XX
AC   AAW53862;
XX
DT   13-JUL-1998  (first entry)
XX
DE   Gravin polypeptide fragment, residues 265-556.
XX
KW   Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW   cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW   Myasthenia gravis; nicotinic acetylcholine receptor.
XX
OS   Homo sapiens.
XX
PN   US5741890-A.
XX
PD   21-APR-1998.
XX
PF   19-DEC-1996;   96US-0769309.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Nauert JB,  Scott JD;
XX
DR   WPI; 1998-260552/23.
XX
PT   New polypeptide fragments of protein kinase binding protein gravin -
PT   are useful for the study of modulation of action between gravin and
PT   protein kinase(s)
XX
PS   Claim 3; Column 17-20; 32pp; English.
XX
CC   This sequence corresponds to residues 265-556 of gravin, and represents
CC   a polypeptide of the invention. The polypeptides are fragments capable of
CC   binding to type II regulatory subunit of cAMP-dependent protein kinase
CC   (PKA). Gravin is a kinase anchoring protein that binds to type II
CC   regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC   antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC   develops antibodies against their own nicotinic acetylcholine receptors.
CC   The polypeptides are useful for providing analogues of gravin in the
CC   study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC   interactions between gravin and kinase. The peptides are involved in the
CC   modulation of gravin-kinase interactions.
XX
SQ   Sequence   292 AA;

  Query Match             11.8%;  Score 952.5;  DB 19;  Length 292;
  Best Local Similarity   66.1%;  Pred. No. 2.7e-43;
  Matches  199;  Conservative   30;  Mismatches   53;  Indels   19;  Gaps    4;

Qy    254 KDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAE 313
Db      1 KEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASE 60

Qy    314 KRKEQEAEKVDEEEKEKTEPASEE---------QEPAEDTDQARLSADYEKVELPLEDQV 364
Db     61 KKKEQEPEKVDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQV 120

Qy    365 GDLEASSEEKCAPLATEVFDEKMEAHQ-EVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEG 423
Db    121 SGSQGPSEEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQ--------KTEVEE 172

Qy    424 TGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIV 483
Db    173 TAGSVPAEELVGMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVV 232

Qy    484 SEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPE 543
Db    233 SEVEMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKR-GGGDEESGEHTQVPADSPD 291

Qy    544 S 544
Db    292 S 292
```

```
RESULT 9
AAB15379
ID   AAB15379 standard; protein; 292 AA.
XX
AC   AAB15379;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Recombinant human gravin polypeptide fragment #1.
XX
KW   Human; gravin; PKA RII binding site; myasthenia gravis;
KW   kinase anchoring protein; cAMP dependent protein kinase.
XX
OS   Homo sapiens.
XX
PN   US6090929-A.
XX
PD   18-JUL-2000.
XX
PF   19-DEC-1997;   97US-0994570.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Scott JD,  Nauert JB;
XX
DR   WPI; 2000-523763/47.
XX
PT   Novel polynucleotides useful for detecting gravin in patients suffering
PT   from Myasthenia gravis encodes cAMP-dependent protein kinase-binding
PT   polypeptide and protein kinase C-binding polypeptide of gravin  -
XX
```

```
PS   Example 4; Column 21-24; 34pp; English.
XX
CC   The present sequence is a recombinant version of the human gravin
CC   protein. Gravin is an A-kinase anchoring protein (AKAP) which is involved
CC   in the localisation of cAMP dependent protein kinase A (PKA) via
CC   interactions between the RII binding region and the PKA regulatory
CC   subunit RII. Gravin is also an antigen found in myasthenia gravis
CC   sufferers, and it is thought that antibodies to it may be useful in
CC   modulating the binding of PKA, and thus aid in the treatment of the
CC   disease. This sequence was used in an assay to determine the ability of
CC   gravin to bind protein kinase C.
XX
SQ   Sequence   292 AA;

  Query Match             11.8%;  Score 952.5;  DB 21;  Length 292;
  Best Local Similarity   66.1%;  Pred. No. 2.7e-43;
  Matches  199;  Conservative   30;  Mismatches   53;  Indels   19;  Gaps    4;

Qy    254 KDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAE 313

Db      1 KEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASE 60

Qy    314 KRKEQEAEKVDEEEKEKTEPASEE---------QEPAEDTDQARLSADYEKVELPLEDQV 364

Db     61 KKKEQEPEKVDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQV 120

Qy    365 GDLEASSEEKCAPLATEVFDEKMEAHQ-EVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEG 423

Db    121 SGSQGPSEEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQ--------KTEVEE 172

Qy    424 TGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIV 483

Db    173 TAGSVPAEELVGMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVV 232

Qy    484 SEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPE 543

Db    233 SEVEMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKR-GGGDEESGEHTQVPADSPD 291

Qy    544 S 544
          |
Db    292 S 292


RESULT 10
AAW53875
ID   AAW53875 standard; peptide; 651 AA.
XX
AC   AAW53875;
XX
DT   13-JUL-1998  (first entry)
XX
DE   Gravin polypeptide fragment, residues 1130-1780.
XX
KW   Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW   cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW   Myasthenia gravis; nicotinic acetylcholine receptor.
XX
OS   Homo sapiens.
XX
PN   US5741890-A.
XX
PD   21-APR-1998.
XX
PF   19-DEC-1996;   96US-0769309.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Nauert JB,  Scott JD;
XX
DR   WPI; 1998-260552/23.
XX
PT   New polypeptide fragments of protein kinase binding protein gravin -
PT   are useful for the study of modulation of action between gravin and
PT   protein kinase(s)
XX
PS   Example 4; Column 49-54; 32pp; English.
XX
CC   This sequence corresponds to residues 1130-1780 of gravin, and represents
CC   a polypeptide of the invention. The polypeptides are fragments capable of
CC   binding to type II regulatory subunit of cAMP-dependent protein kinase
CC   (PKA). Gravin is a kinase anchoring protein that binds to type II
CC   regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC   antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC   develops antibodies against their own nicotinic acetylcholine receptors.
CC   The polypeptides are useful for providing analogues of gravin in the
CC   study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC   interactions between gravin and kinase. The peptides are involved in the
CC   modulation of gravin-kinase interactions.
XX
SQ   Sequence   651 AA;

  Query Match              8.1%;  Score 656;  DB 19;  Length 651;
  Best Local Similarity   38.3%;  Pred. No. 6.1e-27;
  Matches  191;  Conservative   80;  Mismatches  188;  Indels   40;  Gaps   17;

Qy   1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163

Db      6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy   1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222

Db     66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy   1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275

Db    124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy   1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329

Db    182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy   1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386

Db    242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy   1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446

Db    302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy   1447 ESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE--ILELESKSN 1504

Db    362 KKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL---EPENGILELETKSS 416

Qy   1505 KIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRCWTK-----MKDAKM 1558

Db    417 KLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQETEKEGEEPQASAQD 476

Qy   1559 KHPVPQPREDLQVLTVLEA 1577

Db    477 ETPITSAKEESESTAVGQA 495


RESULT 11
AAB18410
ID   AAB18410 standard; protein; 651 AA.
XX
AC   AAB18410;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Human gravin PKA regulatory subunit RII binding site assay peptide #3.
XX
KW   Human; gravin; PKA RII binding site; myasthenia gravis;
KW   kinase anchoring protein; cAMP dependent protein kinase.
```

```
XX
OS   Homo sapiens.
XX
PN   US6090929-A.
XX
PD   18-JUL-2000.
XX
PF   19-DEC-1997;   97US-0994570.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Scott JD,  Nauert JB;
XX
DR   WPI; 2000-523763/47.
XX
PT   Novel polynucleotides useful for detecting gravin in patients suffering
PT   from Myasthenia gravis encodes cAMP-dependent protein kinase-binding
PT   polypeptide and protein kinase C-binding polypeptide of gravin  -
XX
PS   Example 2; Column 53-58; 34pp; English.
XX
CC   The present sequence is part of the human gravin protein. Gravin is an
CC   A-kinase anchoring protein (AKAP) which is involved in the localisation
CC   of cAMP dependent protein kinase A (PKA) via interactions between the
CC   RII binding region and the PKA regulatory subunit RII. Gravin is also an
CC   antigen found in myasthenia gravis sufferers, and it is thought that
CC   antibodies to it may be useful in modulating the binding of PKA, and thus
CC   aid in the treatment of the disease. This sequence was used in an assay
CC   to identify the precise position of the RII binding site within the
CC   gravin protein.
XX
SQ   Sequence   651 AA;

  Query Match             8.1%;  Score 656;  DB 21;  Length 651;
  Best Local Similarity   38.3%;  Pred. No. 6.1e-27;
  Matches  191;  Conservative   80;  Mismatches  188;  Indels   40;  Gaps   17;

Qy   1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
            | |||       :| |:: ||||: |:| || |||||:||||:|| |       ||:|| ::
Db      6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy   1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
          :       |   : || ||  |||| |   |::     |||  |:     | || | |:|::
Db     66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy   1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
           | :|:|||  |   |  |      ||  |       |   : :|    : :|   :|      |    |
Db    124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy   1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
            |:: :: |:||          |   |  ||    ||:|| |  :|   |:||   | |||    |
Db    182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy   1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
          : : |:::|   :   | : || || |       |::|  | ::||||  : :  | || |  |: :
Db    242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy   1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
           ||   | ||||:  |  ||||:  ||||         :          || ||  ||: ||:| |:||:
Db    302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy   1447 ESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE--ILELESKSN 1504
          :   |   ||:||  :  | : :|:| |:    |    :    | | |||   :    |||  |||||:||:
Db    362 KKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL---EPENGILELETKSS 416

Qy   1505 KIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRCWTK-----MKDAKM 1558
          |:| |:|||||||| || ||| |      :  |||      : ||::   :        |           |:
Db    417 KLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQETEKEGEEPQASAQD 476

Qy   1559 KHPVPQPREDLQVLTVLEA 1577
           : |:   :|: :     | :|

Db    477 ETPITSAKEESESTAVGQA 495

RESULT 12
ABB68397
ID   ABB68397 standard; Protein; 2768 AA.
XX
AC   ABB68397;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 31983.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL12500.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 31983; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   2768 AA;

  Query Match             8.0%;  Score 642;  DB 22;  Length 2768;
  Best Local Similarity   22.2%;  Pred. No. 2.6e-25;
  Matches  444;  Conservative  250;  Mismatches  735;  Indels  572;  Gaps   94;

Qy      2 GAGSSTEQRSPEQPAGSDTPSELVLSGHGPA------AEASGAAGDPA-DADPATKLPQK 54
            || |::|: | ||  |  | :   :    ||          ||||  | | :  || :
Db    550 GATSASEE-SEEQDEGKSTEAPTSVDDIEPAKPTESSEEASGEGEDVAKETTPAGE---- 604

Qy     55 NGQLSSVNGVAE--QGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTA 112
              :|: |  |  :|     | : || :||:|     ||     |      :|:   :|
Db    605 ----ASIAGEEEIVKGTTPAGEPSSEG-DEEIVKGTTPAEESSSESE----DELTKVTTP 655

Qy    113 VEDITKDGQEETSEIIEQIPASENNV---EEMVQ---PAESQANDVGFKKVFKFVGFKFT 166
            : :   |:|| ::  |   || |  ::    ||:|:   ||   :::  | :::      ||
Db    656 AGEPSVAGEEEIAK--ETTPAGEPSIAGEEEIVKVTTPA-GESSIAGEEEI-----VKVT 707

Qy    167 VKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGE--SASKESELKQSTE 224
              ::     ::: |    ||    |     || :        |  ||  |  :|  :| :|
Db    708 TPAGESSSEGEEEIIKVTTPAGE----SSSEGDEEIVKESTPAGEPISEGEEDVIKATTS 763
```

```
Qy    225 KQEGTL---KQEQSSTEIPLQAESDQAAE----EEAKDE-------------GEEK---- 260
Db    764 APKSDIEGVKEPETATEVPAEEVEDFAKPTTPIAEAEEEPIAGTPIPTDGISGEEEIVKG 823

Qy    261 -----QEKEPTKSPESPSSPV-----NSETTSSFKKFFTHGWAGWRKKTSFK-------- 302
Db    824 TTPQTLEEQPEISEESTEVPVAEDDLSSSTSASAIASSTEGVQDAASETTSSAPARAGDK 883

Qy    303 ----------KSKEDD--------------------LETAEKRKEQ---EAE------- 321
Db    884 DEAATTVPTAQDKDDEEVEQDATDLPVEDVVQSTTAKTTTTEQPKEESSTEAEDAEIEVT 943

Qy    322 ---KVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVG----DLEA----- 369
Db    944 TSSPADKQEVPEAEPADKDHKDEEDVQTA--------TDLPIKSDIGPPVVDTEATTGQP 995

Qy    370 -SSEEKCA---------PLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQG--------- 410
Db    996 ETSDETATDKPPSVYLPPVSQEVPSSTAKVDNRNDFETEKPTLPPSGEDQSSEPLPAMDL 1055

Qy    411 -GGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAE-----------------ELMKS 452
Db   1056 PAGIPGEGDCLVEG-------KTYANNTIVPATA-PCDVSCKCISSLVACQQMECKLPEN 1107

Qy    453 REMCVSGGD------HTQLTDLSPE-----EKTLPKHPEGIVSEVEMLSSQERIK----- 496
Db   1108 LEKCTVAADLLDGCCPTYICDESTESAEKDEESTAKPDNKIDEDVSEISTEEIPKDVIMP 1167

Qy    497 --VQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSA 554
Db   1168 TGITEQPLSHV-----------------------KPDEEIQPVTSVPAQFDE---STTA 1200

Qy    555 SSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPS 614
Db   1201 KVDKKPIDESAEDKKPI-----GESEEDSKPIDESEED--------KKPVEESAEDKKPV 1247

Qy    615 ESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKP-EEPKRRVDTSVSWEALICVG 673
Db   1248 E-DSEEKEKPLPTVIPASEIEKESKPEDEKKTEADFAAPTEQPEATTPAQIADTAEKEVD 1306

Qy    674 SSKKRARKASSSDDEGGPRTLGGDSHRAEEAS-KDKE--AGTDAVPASTQ---------- 720
Db   1307 DKLATTSAPVSGEDELKP---ADEKKRTETAQIPDAEIPASTDEPESSTELPTVDLDKKP 1363

Qy    721 EQDQAQGSSSPE-------PAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVE 773
Db   1364 EEDSTKGTEAPESDKVPEVPTSASTENE-IEESDKFTTVAPPKISASDETEPTAEEDLV- 1421

Qy    774 QLSTEIEPSREESWVSIKK----------FIPGRRKKR---------AD-GKQEQATVED 813
Db   1422 --PATFEPIESEFEVSTKKPAVQGPPLPTLAPAQPEKKPVDAETSTEADISTEPSAEVEK 1479

Qy    814 SGPVEINEDDPNV-------PAVVPLSEYNAVEREK-MEA-------------QGNTELP 852
Db   1480 EASGETSESDNEIDAGASSTPVPVSADEDKTPSTEKTVEADDKFTTVAPLAGDEEESNLP 1539

Qy    853 QLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEER----------------------- 889
Db   1540 KLPQDIF--EEEAPVAVTTAAPSKDDGEQKPVEVEEKPIEDGQKPIEDETSTPTSSENEI 1597

Qy    890 -----SPSWISASVTEPLEHTAG-----EAMPPVEEVTEKDIIAEETP---VLTQTLPEG 936
Db   1598 EPESDRATTIAPSKEEPSEPSTGAPTKDEPAEPSTDAPESD-ESKETPESEVPTTVAPAG 1656

Qy    937 KDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTE------ASGAEE---TTDMVSA--- 984
Db   1657 EKIPTSSITPDEEPTATSAPVAKPDEDVEKETSTEIPTDAPASSEEDENSSTDQIPSEVP 1716

Qy    985 ---------------VSQLTDSPDTTEEATPVQEV-------------ESGVLDTEE--- 1013
Db   1717 EKKPETPAQTPEEGDIVGATAAPTTSDEVPPVQRLPEEVLAEIPQPSTETGIKQQDETTA 1776

Qy   1014 ---------------EERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEED 1058
Db   1777 APSIDRKEPYVTEIDEEATTVAPISEKDEKPTEE-------------EKPVEQKPTGEEP 1823

Qy   1059 SEVLASEK--EKDVMPKGPVQ-EAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQV 1115
Db   1824 SEEEKEKPIEQDVSTEGPVSTEASEAGSTESSEEVKPSTEG-EVAEKPEDKQPSSTAQA 1882

Qy   1116 IKLQQLMEQAVAPESSETL--TDSETNGSTPLADSDTADGTQQDETIDS------QDSKA 1167
Db   1883 -------PVETIPEISTELPAQDGDKPTSEAPVDSDEDTSAPSDEKIPSVSGEEVEGPEV 1935

Qy   1168 TAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLA 1227
Db   1936 TTASPQA-AEEDELKTPAESEPSS-TDKVPETEYQKPE------DETKADETPESVTQVS 1987

Qy   1228 KTEVGQEGEVDWLDGEK--------VKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
Db   1988 DVATSTSAPVAGGDIEKDEQATTASPEEEEEI----KPTIAPAAEIPQPS---------E 2034

Qy   1280 KESTEVQSLSLEEGEMETDVEKEKR---ETKPEQVSEEGEQETAAPEHEG---------- 1326
Db   2035 KEPVDEQ--EVESGTKATPAESDGQPIDEIAPATSGPIDEASTAAPTKEESTTVASAASP 2092

Qy   1327 ----------TYGKPVL---TLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDT 1373
Db   2093 AVHDDEIKDVTTTQPVADEKEVAAPQDETKTSIDVSTDSPTAQDDEKQDKTEAPV--APT 2150

Qy   1374 TV------------TQTAEAVEKV-IETV----VISETGESPECVGAHLLPAEKSSATGG 1416
Db   2151 TVSSPTADSAADSSTPTVEVPSPVEIDTKPMDDIMSQTIAPHTADGAASTSTEDEDQAPV 2210

Qy   1417 HWTLQHAEDTVPLGPESQ-------------AESIPIIVTPAPE----STLHPDLQGEIS 1459
Db   2211 TVSPQDAEKT-PVSPAPQDSDKTPSSEAPQDADEIPATATPLDDNKIPATVAPQTDDGVP 2269

Qy   1460 ASQRERSEEEDKPDAGP-DADGKESTAI----EKV-LKAEPEILELESKSNKIVLNVIQT 1513
Db   2270 ATAAPLDEDKIQTTAAPLDEEKIPSTAAPLDDEKIPAPVSPVVFDVEPSSEK-------P 2322

Qy   1514 AVDQFARTE-TAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVL 1572
Db   2323 AVSEYDGEESTEPPVHDVETSTDEPT---------------SDAKLKPPTSAPATPSE-S 2366

Qy   1573 TVLEAWAQPRKCLPRLQLKAP 1593
Db   2367 PATEAEIVPETAAPELEKEVP 2387
```

RESULT 13
AAW53867
ID   AAW53867 standard; peptide; 453 AA.
XX
AC   AAW53867;
XX
DT   13-JUL-1998 (first entry)
XX
DE   Gravin polypeptide fragment, residues 1130-1582.
XX
KW   Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW   cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW   Myasthenia gravis; nicotinic acetylcholine receptor.
XX
OS   Homo sapiens.
XX
PN   US5741890-A.
XX
PD   21-APR-1998.
XX
PF   19-DEC-1996;   96US-0769309.
XX
PR   19-DEC-1996;   96US-0769309.
XX

```
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Nauert JB,  Scott JD;
XX
DR   WPI; 1998-260552/23.
XX
PT   New polypeptide fragments of protein kinase binding protein gravin -
PT   are useful for the study of modulation of action between gravin and
PT   protein kinase(s)
XX
PS   Example 2; Column 43-46; 32pp; English.
XX
CC   This sequence corresponds to residues 1130-1582 of gravin, and represents
CC   a polypeptide of the invention. The polypeptides are fragments capable of
CC   binding to type II regulatory subunit of cAMP-dependent protein kinase
CC   (PKA). Gravin is a kinase anchoring protein that binds to type II
CC   regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC   antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC   develops antibodies against their own nicotinic acetylcholine receptors.
CC   The polypeptides are useful for providing analogues of gravin in the
CC   study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC   interactions between gravin and kinase. The peptides are involved in the
CC   modulation of gravin-kinase interactions.
XX
SQ   Sequence   453 AA;

  Query Match             7.9%;  Score 639.5;  DB 19;  Length 453;
  Best Local Similarity   40.5%;  Pred. No. 2.8e-26;
  Matches  183;  Conservative   69;  Mismatches  165;  Indels   35;  Gaps   16;

Qy     1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
            | |||       :| |:: ||||: |:| || |||||:||||:|| |      ||:|| ::
Db        6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy     1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
            :       |    :  || ||  ||||  |      |::    |||  |:     | || | :|::
Db       66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy     1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
             | :|:|||  |  | |     || |       |  |  : :|   : :|  :|    |    |
Db      124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy     1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
             |:: :: |:||          |  | ||    ||:|| | :|  |:||  | |||    |
Db      182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy     1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
            : : |:::|  :   | : || || |      |::| | ::|||| : : | || | |: :
Db      242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy     1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
            ||  | ||||: |   ||||:  ||||       :          || ||  ||: ||:| |:||:
Db      302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy     1447 ESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE--ILELESKSN 1504
            :  |   ||:|| : | : :|:| |:  |   :    | | |||  :   |||   |||||:||:
Db      362 KKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL---EPENGILELETKSS 416

Qy     1505 KIVLNVIQTAVDQFART-ETAPETHAYDSQTQ 1535
            |:| |:||||||||| || ||| |      : |||
Db      417 KLVQNIIQTAVDQFVRTEETATEMLTSELQTQ 448


RESULT 14
AAB18407
ID   AAB18407 standard; protein; 453 AA.
XX
AC   AAB18407;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Human gravin PKA regulatory subunit RII binding site assay peptide #1.
XX
KW   Human; gravin; PKA RII binding site; myasthenia gravis;
KW   kinase anchoring protein; cAMP dependent protein kinase.
XX
OS   Homo sapiens.
XX
PN   US6090929-A.
XX
PD   18-JUL-2000.
XX
PF   19-DEC-1997;  97US-0994570.
XX
PR   19-DEC-1996;  96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Scott JD,  Nauert JB;
XX
DR   WPI; 2000-523763/47.
XX
PT   Novel polynucleotides useful for detecting gravin in patients suffering
PT   from Myasthenia gravis encodes cAMP-dependent protein kinase-binding
PT   polypeptide and protein kinase C-binding polypeptide of gravin  -
XX
PS   Example 2; Column 47-50; 34pp; English.
XX
CC   The present sequence is part of the human gravin protein. Gravin is an
CC   A-kinase anchoring protein (AKAP) which is involved in the localisation
CC   of cAMP dependent protein kinase A (PKA) via interactions between the
CC   RII binding region and the PKA regulatory subunit RII. Gravin is also an
CC   antigen found in myasthenia gravis sufferers, and it is thought that
CC   antibodies to it may be useful in modulating the binding of PKA, and thus
CC   aid in the treatment of the disease. This sequence was used in an assay
CC   to identify the precise position of the RII binding site within the
CC   gravin protein.
XX
SQ   Sequence   453 AA;

  Query Match             7.9%;  Score 639.5;  DB 21;  Length 453;
  Best Local Similarity   40.5%;  Pred. No. 2.8e-26;
  Matches  183;  Conservative   69;  Mismatches  165;  Indels   35;  Gaps   16;

Qy     1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
            | |||       :| |:: ||||: |:| || |||||:||||:|| |      ||:|| ::
Db        6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy     1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
            :       |    :  || ||  ||||  |      |::    |||  |:     | || | :|::
Db       66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy     1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
             | :|:|||  |  | |     || |       |  |  : :|   : :|  :|    |    |
Db      124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy     1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
             |:: :: |:||          |  | ||    ||:|| | :|  |:||  | |||    |
Db      182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy     1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
            : : |:::|  :   | : || || |      |::| | ::|||| : : | || | |: :
Db      242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy     1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
            ||  | ||||: |   ||||:  ||||       :          || ||  ||: ||:| |:||:
Db      302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy     1447 ESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE--ILELESKSN 1504
            :  |   ||:|| : | : :|:| |:  |   :    | | |||  :   |||   |||||:||:
Db      362 KKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL---EPENGILELETKSS 416

Qy     1505 KIVLNVIQTAVDQFART-ETAPETHAYDSQTQ 1535
            |:| |:||||||||| || ||| |      : |||
Db      417 KLVQNIIQTAVDQFVRTEETATEMLTSELQTQ 448
```

```
RESULT 15
ABB66811
ID   ABB66811 standard; Protein; 6815 AA.
XX
AC   ABB66811;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 27225.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10914.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 27225; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   6815 AA;

  Query Match             6.9%;  Score 557.5;  DB 22;  Length 6815;
  Best Local Similarity   20.5%;  Pred. No. 3.1e-20;
  Matches  393;  Conservative  304;  Mismatches  716;  Indels  503;  Gaps   80;

Qy        7 TEQRSPEQPAG----------SDTP-SELVLSGHGPAAEASGAAGDPADADPATKLPQK 54
Db     2989 TTKRMLRRPAGEGEIEIIEVVRDDQPEAEITIVEYEP---------EPVNQDEKPKEPKK 3039

Qy       55 NGQLSSVNGVAEQGDVH--VQ-----------EENQEGQEEEVVDE-------DVGQRE 93
Db     3040 KTR------KVKKDDIHDYIQKLIELETPKTELEKYEKIEFEPIVKDKPLDSPIDVLDES 3093

Qy       94 SEDVREKDR-------------VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEE 140
Db     3094 PKEVQKKDKKSRSTKVPNEETPVQEQYAKVNVVEEEAPEQPEIPVQILEVKPV-EVDVKE 3152

Qy      141 MVQPAESQANDVGFKKVFKFVG------FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEAS 194
Db     3153 VITEDGKPVQEKTTKRVLKKIGPEEQTTFKITM--IESEDNDSV---TVIVDEEPEIASP 3207

Qy      195 VGAGDHQEPSVETAVGESAS-----KESEL-------------KQSTEKQEGTLKQEQSS 236
Db     3208 QSIEEHPEQSKEKLAPKPKKTVRKVKKDDLSDYVKKLIEEEIPKVDLEKYE---KVEMPE 3264

Qy      237 TEIPLQAESDQAAEEEAKDEGE------EKQEKEPTKSPESPSSPVNSETTSSFKKFFTH 290
Db     3265 KPVKLTV-SDSIPEEPKPDKSQPISVLPDTTKPKKTKTPKTPKTEDTDQQVPDEPTETTV 3323

Qy      291 GWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEE---------KEKTEPASEEQEP 340
Db     3324 DTTDIPELTPTQTAQPEDTATAQITPSAQEEKSTQDDTKDTIQKTVKHKKTKP--DTQKS 3381

Qy      341 AEDTDQARLSADY-----------EKVELPLEDQVGDLEASSEEKCAPLATEVFDE--- 385
Db     3382 VETSELPEVHKDYQISIIHEELVEEEQPEKILEVRVID-EVAEVEESQPIVEEVEDEEPQ 3440

Qy      386 ----------------------KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVE 422
Db     3441 PATEETVEDVTKPKSKKKKVVKKKTDDHDELIKKMLEQEIEKTELEKYEKIEFDVPKKLK 3500

Qy      423 GTGESLPPEKLAEPQEVP--------------------QEAEPAEEL--------MKSR 453
Db     3501 PEFAALEPIKIERKEQKPTKVTILDATDVPKTVKLKPSKRKEKPAEELTVQLPKFRLKAR 3560

Qy      454 EMCVSGGDHTQLTDLSPEEKTLPKHPE-GIVSEVEMLS-----SQERIKVQGSPLKKLFS 507
Db     3561 MVLVE----------YPPAPLIPKTTDIGAIKDNGELSRNIEEAEEILKFKPHKTKKI-- 3608

Qy      508 SSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLE 567
Db     3609 ---------KKIK------------------DDLEKVELEKYEKYISSEEEPEEKTPYK 3640

Qy      568 KGPLEAPQDGEAEEGT---TSDGEKK-REGITPWASFKKMVTPKKRVRRPSESDKE---- 619
Db     3641 K-PEKAPKPEEKQEDVKLKLGKGKKKPKEEEAP-----ENVTLKNIPQKPQEVEEEVELK 3694

Qy      620 ---EELEKVKSATLSSTDSTV-------SE------MQDEVKTVGEEQKPEEPKRRVDTS 663
Db     3695 QKPKEVEIVEEQTKKPKDGEFVVEPFEPSEFDRPEYVPDELEQIEHPEIPEKVKKPSKT- 3753

Qy      664 VSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQD 723
Db     3754 -----------KYKPKDKSKSEPETIVSEIVAGVPKEEEAIPEQDVKFRKPERDAPEET 3801

Qy      724 QAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVE-------QLS 776
Db     3802 DSEIKLRPVPQASKDEN------PDEQALVTP--KAEEPIPQEIEDKAIDDEKKPKKSKP 3853

Qy      777 TEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINE-DDPNV----PAVVP 831
Db     3854 KKVQPKEQE----IAKEEP--EEFEVSVKEEEALVDK--PIEIEKPKDVKVKEKKPKEAP 3905

Qy      832 LSEYNAVEREKMEAQGNTELP---QLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTS--V 886
Db     3906 VSEVVVIEEEPKPEEVPEEIPVEYKITTTVLEPEDAPKE--HQVKVIDFDERQETTEEVI 3963

Qy      887 EER----------SPSWISASVTEPLEH------TAGEAMPPVEEVTEKDIIAEETPVL 929
Db     3964 EEKVVTRKKKPKPQQPEEFEVTLKEPKEEQIQPDVVSAEISLPIEEPEQKPEQYEVELKI 4023

Qy      930 TQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQ-- 987
Db     4024 TQTTPE--EPNDVQIAVKEKVKTKPVKKVKEDKIVVVEAEEEKQPVEETIVEVEKQEEKK 4081

Qy      988 LTDSPDTTE-EATPVQEVESGVLDTEEE--ERQTQAILQAVADK-------VKE------ 1031
Db     4082 KSEKPKSYEFKISETQSIEEKPIEVAEEAPEETPKVVEKKVAEKFDSYEFTLKETDEEKV 4141

Qy     1032 -------ESQVPATQTVQRTGSKALEKVEE---------VEEDSEVLASEKEKDVMPKGP 1075
Db     4142 ITVDDQPEEEAP-VEVVFKKKPKEPEAVEAEFVMTEPKIVEETSVETAIKQKKTKKPKKD 4200

Qy     1076 VQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLT 1135
```

```
Db   4201 EEEAQLAIKVVESEAPVAEEVFSEAPE----------SKIVE-----EEVIAEEKPKEFT 4245

Qy   1136 ----DSETNGSTPLADSDTADGTQQDET--------IDSQDSKATAAVRQ---------- 1173
              :||     | :  |    :   |         |  ::||    | :
Db   4246 IRVSESEPKPEEPSVEQFTVKKRKPSVTFADEPATEIVIKESKPAEVVTEDAHIKTKKPK 4305

Qy   1174 SQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKT-EVG 1232
           :||: ||    : :    :|  :|  ||  ||   :|:  |::    |   |  || ::|
Db   4306 KKVTDVEAEELKIKITEEVPQEIPILEEVSEE---EVITETKK----TAPVVEEKTYKIG 4358

Qy   1233 QEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEE 1292
           :         | : ||:|  |             :|  : :|  :           |    | |
Db   4359 IKETEPEKPAEAIVEEEEPVVTEPIEEAPKPEVFEEHKVRVI---------EETPRELVE 4409

Qy   1293 GEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS 1352
            :| :|:  :|:    :: || | |           :  |||           |  :|  |:
Db   4410 EVIEEEVKVIRRKKPKPEIKEEPEAEVTV-----STPKPV--------EEVEATSSIAVI 4456

Qy   1353 PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISE--------------TGESP 1398
          |  | :::|  ::: :    :    |    | |:::  |  :: |                | :|
Db   4457 PEQPTEEEAADLKITIIEEE---TPPQELVQEIEEIEIVEEPKAPEEQPTDFTFATKDSE 4513

Qy   1399 ECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG---PESQAESIPIIVTPAPESTLHPDLQ 1455
          :         || |:           |:|   :     |:       | :||  :      ||  | |         :
Db   4514 KKPTVEELPEEQV-------TIQKKKKAPVPEVVEEPEAEFLVKPKTPVQEVT----EE 4562

Qy   1456 GEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILEL--------ESKSNKIV 1507
           :|: |::    |||        |    :      |  : :   |||: |:          |  |   : |
Db   4563 AKITKSKKPVKEEE--------AAAELKVTITEEIPTEPEVQEIIEEIEEIEEEKPAEYV 4614

Qy   1508 LNVIQTAVDQFARTET--------APETHAYDSQ-TQVPACRLDSREPNRCWTKMKDAKM 1558
          : | :: :      |          ||        ::: |   |   : : :        | |  |:
Db   4615 IEVKESQPEAVEDKEVSLPKKKPKAPIVEEPEAEITLKPKVKSEEVQEEAKIVKKKPKKI 4674

Qy   1559 -------------KHPVPQPREDLQVLTVLEAWAQPRKCLPR-------LQLKAP 1593
                        :  ||:|  : :|:    |     :|::     |          ::|| |
Db   4675 DEVAVADELTVKVEEEVVPEPIVEEEVIEEFEIKKKPKEPEPEDIVDAAIVKLKKP 4730
```

Search completed: December 13, 2002, 00:12:16
Job time : 141 secs

```
                 GenCore version 5.1.3
           Copyright (c) 1993 - 2002  Compugen Ltd.

OM protein - protein search, using sw model

Run on:          December 13, 2002, 00:10:50 ; Search time 38 Seconds
                                          (without alignments)
                                          1235.763 Million cell updates/sec

Title:           US-09-902-432-4
Perfect score:   8073
Sequence:        1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters:      262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Issued_Patents_AA:*
                 1:  /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2:  /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3:  /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4:  /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5:  /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6:  /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5421.5	67.2	1346	2	US-08-635-121-2	Sequence 2, Appli
2	4274.5	52.9	1780	1	US-08-769-309A-5	Sequence 5, Appli
3	4274.5	52.9	1780	3	US-08-994-570-5	Sequence 5, Appli
4	952.5	11.8	292	1	US-08-769-309A-3	Sequence 3, Appli
5	952.5	11.8	292	3	US-08-994-570-3	Sequence 3, Appli
6	656	8.1	651	1	US-08-769-309A-17	Sequence 17, Appl
7	656	8.1	651	3	US-08-994-570-17	Sequence 17, Appl
8	639.5	7.9	453	1	US-08-769-309A-14	Sequence 14, Appl
9	639.5	7.9	453	3	US-08-994-570-14	Sequence 14, Appl
10	521.5	6.5	396	1	US-08-769-309A-15	Sequence 15, Appl
11	521.5	6.5	396	3	US-08-994-570-15	Sequence 15, Appl
12	462	5.7	1786	4	US-08-973-462-8	Sequence 8, Appli
13	388	4.8	1805	1	US-07-853-913-2	Sequence 2, Appli
14	365.5	4.5	3696	4	US-09-134-001C-5080	Sequence 5080, Ap
15	361.5	4.5	2101	1	US-08-466-390-4	Sequence 4, Appli
16	361.5	4.5	2101	1	US-08-470-950-4	Sequence 4, Appli
17	361.5	4.5	2101	1	US-08-467-781-4	Sequence 4, Appli
18	361.5	4.5	2101	2	US-08-483-924-4	Sequence 4, Appli
19	361.5	4.5	2101	4	US-09-452-294-1	Sequence 1, Appli
20	361	4.5	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
21	358.5	4.4	2101	1	US-08-195-487-4	Sequence 4, Appli
22	358.5	4.4	2101	5	PCT-US93-06160-4	Sequence 4, Appli
23	354	4.4	1618	1	US-07-853-913-4	Sequence 4, Appli
24	354	4.4	2409	6	5180808-2	Patent No. 5180808
25	334.5	4.1	1162	2	US-08-728-323A-2	Sequence 2, Appli
26	334.5	4.1	1162	4	US-09-298-568-2	Sequence 2, Appli
27	334	4.1	1018	1	US-08-072-610-2	Sequence 2, Appli
28	334	4.1	1018	2	US-08-719-822B-2	Sequence 2, Appli
29	334	4.1	1018	4	US-09-092-458-2	Sequence 2, Appli
30	321	4.0	1852	1	US-08-425-061-24	Sequence 24, Appl
31	321	4.0	1852	2	US-08-825-886-24	Sequence 24, Appl
32	321	4.0	1863	1	US-08-425-061-16	Sequence 16, Appl
33	321	4.0	1863	1	US-08-480-784-2	Sequence 2, Appli
34	321	4.0	1863	1	US-08-483-553-2	Sequence 2, Appli
35	321	4.0	1863	1	US-08-487-002-2	Sequence 2, Appli
36	321	4.0	1863	1	US-08-483-554B-2	Sequence 2, Appli
37	321	4.0	1863	1	US-08-488-011B-2	Sequence 2, Appli
38	321	4.0	1863	2	US-08-825-886-16	Sequence 16, Appl
39	321	4.0	1863	4	US-08-850-727-2	Sequence 2, Appli
40	321	4.0	1863	5	PCT-US95-10202-2	Sequence 2, Appli
41	321	4.0	1863	5	PCT-US95-10203-2	Sequence 2, Appli
42	321	4.0	1863	5	PCT-US95-10220-2	Sequence 2, Appli
43	320	4.0	1898	1	US-08-056-200-94	Sequence 94, Appl
44	320	4.0	1898	2	US-08-800-644-94	Sequence 94, Appl
45	319.5	4.0	1939	4	US-09-310-187A-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-635-121-2
; Sequence 2, Application US/08635121
; Patent No. 5910442
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,121
;      FILING DATE:  19-APRIL-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/33603
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1346 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  peptide
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  internal
;    ORIGINAL SOURCE:
US-08-635-121-2

  Query Match             67.2%;  Score 5421.5;  DB 2;  Length 1346;
  Best Local Similarity   91.6%;  Pred. No. 0;
```

```
  Matches 1098;  Conservative   24;  Mismatches   62;  Indels   15;  Gaps    7;

Qy        387 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 60

Qy        447 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 120

Qy        507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 180

Qy        567 EKGPLEAPQDGEAEEGTTSDGEKKRE----GITPWASFKKMVTPKKRVRRPSESDKEEEL 622
              |||||||: |:  :      |||||:      ||      :|      :: ||||||||||||
Db        181 EKGPLEAPRMGKLRKELLR-GEKKRKDHSLGI-----LQKDGDTQETVRRPSESDKEEEL 234

Qy        623 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 294

Qy        683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 742
              ||||   ||||||    ::     : : : |||||||||||||||||||||||||||||
Db        295 SSSDIR-GPRTLGGGQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 353

Qy        743 VSTWESFKRLVTPRKKSKSKLEEK-AEDSSVEQLSTEIEPSREESWVS--IKKFIPGRRK 799
              ||||||||||||||||||||||| |  : |          :  :| |   :   ::    |   :
Db        354 VSTWESFKRLVTPRKKSKSKLEEKEAGRTLVVGAGCPLRSNRVEKNLGFPLRNSSPDGGR 413

Qy        800 KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVY 859
              |    |:|||||||||||||||:|:||||||||:|||||||||||| |||    |  |
Db        414 KGQMGRQEQATVEDSGPVEINEDEPDVPAVVPLSEYDAVEREKMEAQGNAELPSCWGCV- 472

Qy        860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        473 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 532

Qy        920 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 979
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        533 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 592

Qy        980 DMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039
              |||||||||||||||||||||||| |||||||||||||||||||||||||||||||||||
Db        593 DMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 652

Qy       1040 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 1099
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        653 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 712

Qy       1100 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 1159
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        713 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 772

Qy       1160 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELT 1219
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        773 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELA 832

Qy       1220 AAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
              |||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        833 AAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 892

Qy       1280 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPS 1339
              ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||||
Db        893 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVLTLDMPS 952

Qy       1340 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        953 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1012

Qy       1400 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1459
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1013 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1072

Qy       1460 ASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1519
              |||||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db       1073 ASQRERSEEEDKPDAGPDADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1132

Qy       1520 RTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAW 1578
              ||||||||||||||||||||  |||||||||||||| |||||||||||||||||||||||
Db       1133 RTETAPETHAYDSQTQVPAMRLDSREPNRCWTKMKVAKMKHPVPQPREDLQVLTVLEAW 1191


RESULT 2
US-08-769-309A-5
; Sequence 5, Application US/08769309A
; Patent No. 5741890
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,309A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5741890and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1780 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-769-309A-5

  Query Match             52.9%;  Score 4274.5;  DB 1;  Length 1780;
  Best Local Similarity   57.8%;  Pred. No. 9.2e-255;
  Matches  957;  Conservative  187;  Mismatches  402;  Indels  109;  Gaps   36;

Qy          1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
              ||||||||||||||  |  || |:|   ||  ||:|||:       |||  |    ||||| ||||
Db          1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy         57 QLSSVNGVAEQGDVHVQEENQEGQ-------------EEEVVDEDVGQRESEDVREKDR 102
              |||:|||||  :: :| :  |||             ||||:   :||||:||||  |:|
Db         61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy        103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
                :|||  |  |  |||  |||||     |||||:||:|:|:||: || ||||||:||||||||||
Db        121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy        163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
              |||||||| || ||||||||||||||         ||||||:||:        ||:|||||| |||
Db        180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy        223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
```

```
Db     235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy     283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337

Db     294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy     338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392

Db     354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy     393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452

Db     414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy     453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512

Db     466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy     513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572

Db     526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy     573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632

Db     585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy     633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692

Db     645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPK 704

Qy     693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752

Db     705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy     753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808

Db     765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy     809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865

Db     825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy     866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924

Db     885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy     925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984

Db     945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy     985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038

Db    1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy    1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097

Db    1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy    1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147

Db    1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy    1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207

Db    1179 FDAPGTTQKDEIVEIHEENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1236

Qy    1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259

Db    1237 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1294

Qy    1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314

Db    1295 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1354

Qy    1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372

Db    1355 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1414

Qy    1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430

Db    1415 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1474

Qy    1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490

Db    1475 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1531

Qy    1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547

Db    1532 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1589

Qy    1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577

Db    1590 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1624
```

```
RESULT 3
US-08-994-570-5
; Sequence 5, Application US/08994570
; Patent No. 6090929
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,570
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090929and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  5:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1780 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-994-570-5

  Query Match            52.9%;  Score 4274.5;  DB 3;  Length 1780;
  Best Local Similarity  57.8%;  Pred. No. 9.2e-255;
  Matches  957;  Conservative  187;  Mismatches  402;  Indels  109;  Gaps   36;

Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56

Db       1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60
```

```
Qy     57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102
Db     61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy    103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db    121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy    163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db    180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy    223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db    235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy    283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db    294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy    338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db    354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy    393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db    414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy    453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db    466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy    513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db    526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy    573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db    585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy    633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db    645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPK 704

Qy    693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db    705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy    753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
Db    765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy    809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db    825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db    885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy    925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db    945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy    985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db   1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy   1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db   1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy   1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db   1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy   1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db   1179 FDAPGTTQKDEIVEIHEENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1236

Qy   1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db   1237 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1294

Qy   1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db   1295 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1354

Qy   1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db   1355 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1414

Qy   1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db   1415 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1474

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db   1475 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1531

Qy   1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547
Db   1532 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1589

Qy   1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db   1590 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1624
```

RESULT 4
US-08-769-309A-3
; Sequence 3, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/769,309A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5741890and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 292 amino acids

```
;       TYPE:  amino acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
US-08-769-309A-3

  Query Match             11.8%;  Score 952.5;  DB 1;  Length 292;
  Best Local Similarity   66.1%;  Pred. No. 2.5e-51;
  Matches  199;  Conservative   30;  Mismatches   53;  Indels   19;  Gaps    4;

Qy    254 KDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAE 313
          |:|||||||||:|| |||:||| ||| |:|||||| |||||||||||:| |||::| :|
Db      1 KEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASE 60

Qy    314 KRKEQEAEKVDEEEKEKTEPASEE---------QEPAEDTDQARLSADYEKVELPLEDQV 364
          |:|||| |||| ||   | | |||:         |||||    : ||||:|||||| |:||
Db     61 KKKEQEPEKVDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQV 120

Qy    365 GDLEASSEEKCAPLATEVFDEKMEAHQ-EVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEG 423
             :  |||| |||||||||||:| || |||||||||||:   |||          ||
Db    121 SGSQGPSEEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQ--------KTEVEE 172

Qy    424 TGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIV 483
          |  |:| |:|          |||||||:||:| :| |||| | ||  ||||:|| | | |||:|
Db    173 TAGSVPAEELVGMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVV 232

Qy    484 SEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPE 543
          ||||||||||:|||||||||||:|:||||||||||||| ||||| ||:  :  :||:
Db    233 SEVEMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKR-GGGDEESGEHTQVPADSPD 291

Qy    544 S 544
          |
Db    292 S 292

RESULT 5
US-08-994-570-3
; Sequence 3, Application US/08994570
; Patent No. 6090929
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,570
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090929and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  292 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-994-570-3

  Query Match             11.8%;  Score 952.5;  DB 3;  Length 292;
  Best Local Similarity   66.1%;  Pred. No. 2.5e-51;
  Matches  199;  Conservative   30;  Mismatches   53;  Indels   19;  Gaps    4;

Qy    254 KDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAE 313
          |:|||||||||:|| |||:||| ||| |:|||||| |||||||||||:| |||::| :|
Db      1 KEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASE 60

Qy    314 KRKEQEAEKVDEEEKEKTEPASEE---------QEPAEDTDQARLSADYEKVELPLEDQV 364
          |:|||| |||| ||   | | |||:         |||||    : ||||:|||||| |:||
Db     61 KKKEQEPEKVDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQV 120

Qy    365 GDLEASSEEKCAPLATEVFDEKMEAHQ-EVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEG 423
             :  |||| |||||||||||:| || |||||||||||:   |||          ||
Db    121 SGSQGPSEEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQ--------KTEVEE 172

Qy    424 TGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIV 483
          |  |:| |:|          |||||||:||:| :| |||| | ||  ||||:|| | | |||:|
Db    173 TAGSVPAEELVGMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVV 232

Qy    484 SEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPE 543
          ||||||||||:|||||||||||:|:||||||||||||| ||||| ||:  :  :||:
Db    233 SEVEMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKR-GGGDEESGEHTQVPADSPD 291

Qy    544 S 544
          |
Db    292 S 292

RESULT 6
US-08-769-309A-17
; Sequence 17, Application US/08769309A
; Patent No. 5741890
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,309A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5741890and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  17:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  651 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-769-309A-17
```

```
  Query Match              8.1%;  Score 656;  DB 1;  Length 651;
  Best Local Similarity   38.3%;  Pred. No. 1.4e-32;
  Matches  191;  Conservative   80;  Mismatches  188;  Indels   40;  Gaps   17;

Qy       1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
              | |||       :| |:: ||||: |:| || |||||:||||:|| |      ||:|| ::
Db          6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy       1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
              :       |    : || ||  |||| |    |::    |||  |:      | || | :|::
Db         66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy       1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
              | :|:|||  |  | |     || |       |  |  : :|   : :|  :|     |    |
Db        124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy       1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
               |:: :: |:||        |  | ||   ||:|| | :|  |:||  | |||       |
Db        182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy       1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
              : : |:::|  :   |  : || ||  |      |::| | ::|||| : :  | || | |: :
Db        242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy       1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
              ||  | ||||: |   ||||:  ||||           :        || || ||: ||:| |:||:
Db        302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy       1447 ESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE--ILELESKSN 1504
               :  |  ||:|| : | : :|:| |:  |   :    | | |||  :    |||  |||||:||:
Db        362 KKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL---EPENGILELETKSS 416

Qy       1505 KIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRCWTK-----MKDAKM 1558
              |:| |:||||||| || |||  |       : |||       : ||::   :     |        |:
Db        417 KLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQETEKEGEEPQASAQD 476

Qy       1559 KHPVPQPREDLQVLTVLEA 1577
               : |:    :|: :    | :|
Db        477 ETPITSAKEESESTAVGQA 495
```

RESULT 7
US-08-994-570-17
; Sequence 17, Application US/08994570
; Patent No. 6090929
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/994,570
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090929and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 651 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-994-570-17

```
  Query Match              8.1%;  Score 656;  DB 3;  Length 651;
  Best Local Similarity   38.3%;  Pred. No. 1.4e-32;
  Matches  191;  Conservative   80;  Mismatches  188;  Indels   40;  Gaps   17;

Qy       1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
              | |||       :| |:: ||||: |:| || |||||:||||:|| |      ||:|| ::
Db          6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy       1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
              :       |    : || ||  |||| |    |::    |||  |:      | || | :|::
Db         66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy       1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
              | :|:|||  |  | |     || |       |  |  : :|   : :|  :|     |    |
Db        124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy       1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
               |:: :: |:||        |  | ||   ||:|| | :|  |:||  | |||       |
Db        182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy       1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
              : : |:::|  :   |  : || ||  |      |::| | ::|||| : :  | || | |: :
Db        242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy       1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
              ||  | ||||: |   ||||:  ||||           :        || || ||: ||:| |:||:
Db        302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy       1447 ESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE--ILELESKSN 1504
               :  |  ||:|| : | : :|:| |:  |   :    | | |||  :    |||  |||||:||:
Db        362 KKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL---EPENGILELETKSS 416

Qy       1505 KIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPNRCWTK-----MKDAKM 1558
              |:| |:||||||| || |||  |       : |||       : ||::   :     |        |:
Db        417 KLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQETEKEGEEPQASAQD 476

Qy       1559 KHPVPQPREDLQVLTVLEA 1577
               : |:    :|: :    | :|
Db        477 ETPITSAKEESESTAVGQA 495
```

RESULT 8
US-08-769-309A-14
; Sequence 14, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```
;     SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,309A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5741890and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  453 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-769-309A-14

  Query Match             7.9%;  Score 639.5;  DB 1;  Length 453;
  Best Local Similarity   40.5%;  Pred. No. 8.7e-32;
  Matches  183;  Conservative   69;  Mismatches  165;  Indels   35;  Gaps   16;

Qy    1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
Db       6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy    1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
Db      66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy    1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
Db     124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy    1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
Db     182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy    1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
Db     242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy    1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
Db     302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy    1447 ESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE--ILELESKSN 1504
Db     362 KKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL---EPENGILELETKSS 416

Qy    1505 KIVLNVIQTAVDQFART-ETAPETHAYDSQTQ 1535
Db     417 KLVQNIIQTAVDQFVRTEETATEMLTSELQTQ 448
```

RESULT 9

```
US-08-994-570-14
; Sequence 14, Application US/08994570
; Patent No. 6090929
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,570
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090929and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  14:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  453 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-994-570-14

  Query Match             7.9%;  Score 639.5;  DB 3;  Length 453;
  Best Local Similarity   40.5%;  Pred. No. 8.7e-32;
  Matches  183;  Conservative   69;  Mismatches  165;  Indels   35;  Gaps   16;

Qy    1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
Db       6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy    1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
Db      66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy    1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
Db     124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy    1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
Db     182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy    1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
Db     242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy    1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
Db     302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy    1447 ESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE--ILELESKSN 1504
Db     362 KKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL---EPENGILELETKSS 416

Qy    1505 KIVLNVIQTAVDQFART-ETAPETHAYDSQTQ 1535
Db     417 KLVQNIIQTAVDQFVRTEETATEMLTSELQTQ 448
```

RESULT 10

```
US-08-769-309A-15
; Sequence 15, Application US/08769309A
; Patent No. 5741890
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
```

```
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,309A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5741890and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  396 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-769-309A-15

  Query Match             6.5%;  Score 521.5;  DB 1;  Length 396;
  Best Local Similarity   38.2%;  Pred. No. 1.3e-24;
  Matches  147;  Conservative   62;  Mismatches  149;  Indels   27;  Gaps   12;

Qy   1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
          | |||       :| |:: ||||: |:| || |||||:||||:|| |      ||:|| ::
Db      6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy   1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
          :        |   : || ||  |||| |     |::   |||  |:      | || | :|::
Db     66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy   1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
          | :|:|||  |   |  |      ||  |     |  |  : :|    : :|   :|    |    |
Db    124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy   1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
           |:: :: |:||       |   |  ||    ||:|| | :|   |:||  | |||     |
Db    182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy   1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
          : : |:::|  :   | : || || |      |::|  | ::|||| : :  | || | |: :
Db    242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy   1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
          ||  | ||||: |  ||||:  ||||        :       || || ||: ||:| |:||:
Db    302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy   1447 ESTLHPDLQGEISASQRERSEEEDK 1471
          :  |  ||:|| : | : :|:| |:
Db    362 KKGLSSDLEGEKTTSLKWKSDEVDE 386

RESULT 11
US-08-994-570-15
; Sequence 15, Application US/08994570
; Patent No. 6090929
;  GENERAL INFORMATION:
;    APPLICANT:  Scott, John D.,
;    APPLICANT:  Nauert, Brian J.,
;    APPLICANT:  Klauck, Theresa M.
;    TITLE OF INVENTION:  Protein Binding Domains of Gravin
;    NUMBER OF SEQUENCES:  24
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Marshall, O'Toole, Gerstein, Murray & Borun
;      STREET:  6300 Sears Tower/233 South Wacker Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/994,570
;      FILING DATE:
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 6090929and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  15:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  396 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
US-08-994-570-15

  Query Match             6.5%;  Score 521.5;  DB 3;  Length 396;
  Best Local Similarity   38.2%;  Pred. No. 1.3e-24;
  Matches  147;  Conservative   62;  Mismatches  149;  Indels   27;  Gaps   12;

Qy   1110 VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQ 1163
          | |||       :| |:: ||||: |:| || |||||:||||:|| |      ||:|| ::
Db      6 VTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIH 65

Qy   1164 DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG-RDVLEPTQQELTAAA 1222
          :        |   : || ||  |||| |     |::   |||  |:      | || | :|::
Db     66 EENEVHLV-PVRGTEAEAVPAQKERPPA-PSSFVFQEETKEQSKMEDTLEHTDKEVSVET 123

Qy   1223 VPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNSQKAADVTYDSEVMGVA 1275
          | :|:|||  |   |  |      ||  |     |  |  : :|    : :|   :|    |    |
Db    124 VSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEA 181

Qy   1276 GCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEEG-EQETAAPEHEGTYG 1329
           |:: :: |:||       |   |  ||    ||:|| | :|   |:||  | |||     |
Db    182 ECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSK 241

Qy   1330 KPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVI-- 1386
          : : |:::|  :   | : || || |      |::|  | ::|||| : :  | || | |: :
Db    242 QLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLG 301

Qy   1387 ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAP 1446
          ||  | ||||: |  ||||:  ||||        :       || || ||: ||:| |:||:
Db    302 ETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATT 361

Qy   1447 ESTLHPDLQGEISASQRERSEEEDK 1471
          :  |  ||:|| : | : :|:| |:
Db    362 KKGLSSDLEGEKTTSLKWKSDEVDE 386

RESULT 12
US-08-973-462-8
; Sequence 8, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
```

```
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
;   LENGTH: 1786
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: Description of Artificial Sequence:Polypeptide
US-08-973-462-8

  Query Match             5.7%;  Score 462;  DB 4;  Length 1786;
  Best Local Similarity   19.2%;  Pred. No. 5e-20;
  Matches  325;  Conservative  291;  Mismatches  644;  Indels  436;  Gaps   69;

Qy     53 QKNGQLSSVNGVAEQGDVHVQEENQEGQE-EEVVDEDVGQRESEDVREK---DRVEE--- 105
Db     95 QVNGELAS-----EEVKEKILDLLEEGNTLTESVDDNKNLEEAEDIKENILLSNIEEPKE 149

Qy    106 ---------MAANSTAVEDITKDGQ---EETSEIIEQIPASENNVEEMVQPAESQANDVG 153
Db    150 NIIDNLLNNIGQNSEKQESVSENVQVSDELFNELLNSVDVNGEVKENILE--ESQVNDDI 207

Qy    154 FKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASV--GAGDHQEPSVETAVGE 211
Db    208 FNSLVK------SVQQEQQHNVEEKVEESVEENDEESVEENVEENVEENDDGSVASSVEE 261

Qy    212 S--ASKESELKQSTEKQEGTLKQE------------------QSSTEIPLQAESDQAAEE 251
Db    262 SIASSVDESIDSSIEENVAPTVEEIVAPSVVESVAPSVEESVEENVEESVAENVEESVAE 321

Qy    252 EAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLE- 310
Db    322 NVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVEESVEENVEE 381

Qy    311 -TAEKRKEQEAEKVDEEEKEKTEPA---SEEQEPAEDTDQARLSADYEKVELPLEDQVGD 366
Db    382 SVAENVEESVAENVEESVAENVEESVAENVEESVAENVEEI-VAPTVEEIVAPTVEEI-- 438

Qy    367 LEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGE 426
Db    439 VAPSVVESVAPSVEESVEENVE--ESVAENVEESVAENVEES-----------VAENVEE 485

Qy    427 SLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEV 486
Db    486 SV--------------AENVEESVAE----------------NVEESVAENVEESVAENV 515

Qy    487 EMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESAD 546
Db    516 EEIVAPTVEEIVAPTVEEIVAPSVVESVA---------------PSVEESVEENVEESVA 560

Qy    547 EQKGESSASSPEEP-----EETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFK 601
Db    561 ENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTV------EEIVAP--SVV 612

Qy    602 KMVTPKKRVRRPSESDKEEEL-EKVKSATLSSTDSTVSEMQDEV--KTVGEEQKPEEPKR 658
Db    613 ESVAPS--VEESVEENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP----- 665

Qy    659 RVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEE--ASKDKEAGTDAVP 716
Db    666 ------TVEEIVAPSVVESVAPSVEESVEE-----------NVEESVAENVEESVAENVE 708

Qy    717 ASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTP------RKKSKSKLEEKAEDS 770
Db    709 ESVAENVEESVAENVEEIVAPTVEEIVA--PTVEEIVAPSVVESVAPSVEESVEENVEES 766

Qy    771 SVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQA--TVEDSGPVEINEDDPNVPA 828
Db    767 VAENVEESVAENVEES-------------VAENVEESVAPTVEEIVAPSVEE------S 806

Qy    829 VVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEE 888
Db    807 VAPSVEESVAENVATNLSDNL-LSNLLGGI-ETEEIKDSILNEIE----------EVKE 853

Qy    889 RSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEG-KDAHDDMVTSE 947
Db    854 N----VVTTILENVEETTAESVTTFSNILEE---IQENTITNDTIEEKLEELHENVLSAA 906

Qy    948 VDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQ--------LTDSPDTTEEAT 999
Db    907 LENTQSEEEKKEVIDVI--EEVKE----EVATTLIETVEQAEEKSANTITEIFENLEE-- 958

Qy   1000 PVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKA-----LEKVEE 1054
Db    959 --NAVESNENVAENLEKLNETVFNTVLDKVEETVEISGESLENNEMDKAFFSEIFDNVKG 1016

Qy   1055 VEED-----------SEVLASEKEKDV--------------MPKG------------PVQ 1077
Db   1017 IQENLLTGMFRSIETSIVIQSEEKVDLNENVVSSILDNIENMKEGLLNKLENISSTEGVQ 1076

Qy   1078 EAGAEHLAQG------------------SETGQATPESLEVPEVTADVDHVATCQVIK-- 1117
Db   1077 ETVTEHVEQNVYVDVDVPAMKDQFLGILNEAGGLKEMFFNLEDVFKSESDVITVEEIKDE 1136

Qy   1118 -LQQLMEQ---AVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQ 1173
Db   1137 PVQKEVEKETVSIIEEMEENIVDVLEEEKEDLTDK-MIDAVEESIEISSDSKEETESIKD 1195

Qy   1174 SQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLE--PTQQELTAAAVPV------ 1225
Db   1196 K---EKDVSLVVEE----------VQDNDMDESVEKVLELKNMEEELMKDAVEINDITSK 1242

Qy   1226 -LAKTEVGQEGEVDWL-DGEKVKE---------EQEVFVHSGPNSQKAADVTYD-----SE 1270
Db   1243 LIEETQELNEVEADLIKDMEKLKELEKALSEDSKEIIDAKDDTLEKVIEEEHDITTTLDE 1302

Qy   1271 VMGVAGCQE--------------------KESTEVQSLSLEEGE----METDVEKEKRET 1306
Db   1303 VVELKDVEEDKIEKVSDLKDLEEDILKEVKEIKELESEILEDYKELKTIETDILEEKKEI 1362

Qy   1307 KPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSE------------------------- 1341
Db   1363 EKDHFEKFEEEAEEIKDLEADILKEVSSLEVEEEKKLEEVHELKEEVEHIISGDAHIKGL 1422

Qy   1342 -----------RGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVV 1390
Db   1423 EEDDLEEVDDLKGSILDMLKGDMELGDMDK---------ESLEDVTTKLGERVES-LKDVL 1473

Qy   1391 ISETGESPECVGAH-----------LL-------PAEKSSATGGHWTLQHAE--DTVPLG 1430
Db   1474 SSALGMDEEQMKTRKKAQRPKLEEVLLKEEVKEEPKKKITKKKVRFDIKDKEPKDEI-VE 1532

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db   1533 VEMKDEDI---------EEDVEEDIEEDIEEDKVEDIDEDIDEDIGEDKD----EVIDLIV 1580

Qy   1491 KAEPEILELESKSNKI 1506
Db   1581 QKEKRIEKVKAKKKKL 1596

RESULT 13
US-07-853-913-2
; Sequence 2, Application US/07853913
; Patent No. 5338839
;  GENERAL INFORMATION:
```

```
;    APPLICANT:  McKay, Ronald D.G.
;    APPLICANT:  Lendahl, Urban
;    TITLE OF INVENTION:  Nestin Expression As An Indicator of
;    TITLE OF INVENTION:  Neuroepithelial Tumors
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Hamilton, Brook, Smith & Reynolds, P.C.
;      STREET:  Two Militia Drive
;      CITY:  Lexington
;      STATE:  Massachusetts
;      COUNTRY:  U.S.A.
;      ZIP:  02173
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/07/853,913
;      FILING DATE:  19920319
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/660,412
;      FILING DATE:  22-FEB-1991
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/603,803
;      FILING DATE:  25-OCT-1990
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/201,762
;      FILING DATE:  02-JUN-1988
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  US 07/180,548
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Granahan, Patricia
;      REGISTRATION NUMBER:  32,227
;      REFERENCE/DOCKET NUMBER:  MIT-4641AAAA
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  617-861-6240
;      TELEFAX:  617-861-9540
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1805 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-07-853-913-2

  Query Match             4.8%;  Score 388;  DB 1;  Length 1805;
  Best Local Similarity   21.3%;  Pred. No. 1.8e-15;
  Matches  319;  Conservative  233;  Mismatches  543;  Indels  402;  Gaps   75;

Qy     13 EQPAGSDT---PSELV-----LSGHGPAAEASGAAGDPADADPATKLPQKNGQLSSVNGV 64
Db    440 EEPGGKQQGHFPDDLTSLATNLNPHHPTLEAKD--GESSESRVSSIFQEDEGQIWEL--V 495

Qy     65 AEQGDVHVQEENQEGQ--EEEVVDEDVGQRESEDVREKDRVEEMAAN---STAVEDITKD 119
Db    496 EKEADIEVKVENSSAQKTQESGLDTEETQ-DSQGPLQKETLKALGEEPLMSLKIQNYETA 554

Qy    120 GQE------------ETSEIIEQIP---ASENNVEEMVQPAESQANDVGFKKVFKFVGF 163
Db    555 GKENCNSSTEGHLGTLEGPEKEKQIPLKSLEEKNVE------SEKTLENGVPVLSELLGK 608

Qy    164 KFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVET----------AVGESA 213
Db    609 EDTRTEDQELMSPKG---TLKRFSSLGKESQEVVRPSKEGNLESWTAFKKEESQHPLGFPG 665

Qy    214 SKESELKQSTEKQEGTL----KQEQSSTEIPLQAESDQ-AAEEEAKDEGE------EKQE 262
Db    666 AEDQMLERLVEKEDQSFPRSPEEEDQEACRPLQKENQEPLGYEEA--EGQILERLIEKES 723

Qy    263 KEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEK 322
Db    724 QESLRSPEEE----DQEAGRSLQK-------ENQEPLGYEEAEDQMLERLIEKESQESLK 772

Qy    323 VDEEEKEKTEPASEEQE-------------PAEDTDQARLSA-----DYEKVELPLE-- 361
Db    773 SPEENQRIGKPLERENQKSLRYLEENQETFVPLESRNQRPLRSLEVEEEEQRIVKPLEKV 832

Qy    362 --DQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGV 419
Db    833 SQDSLGSL---AEENVQPLR---YLEEDDCINKSLLE------DKTHKSLGSLEDRNGDS 880

Qy    420 VV----EGTGESL-PPEK---------LAEPQEVPQEAEPAEELMKSREMCVSGGDHTQL 465
Db    881 IIIPQESETQVSLRPPEEEDQRIVNHLEKESQEFSRSSEEEEQVM---ERSLEGENHESL 937

Qy    466 TDLSPE----EKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKG 521
Db    938 SSVEKEDQMVESQLEKESQDSGKSLED-ESQETF----GPLEK-ENAESLRSLAGQDQ-- 989

Qy    522 KRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQD-GEAE 580
Db    990 ----------EEQKLEQETQQTLRAVGNEQMAVSPP---------EKVDPELPKPLGNDQ 1030

Qy    581 EGTTSDGEKKREGITPWASFKKMVTPKK---RVRRPSESDKEEELEKVKS----ATLSST 633
Db   1031 EIARSLGKENQESLVSLKE-KGIETVKSLETEIIEPLET-AEEDLERRKSIDTQEPLWST 1088

Qy    634 D---STVSEMQDE----VKTVGEEQKP----EEPKRRVDTSVSWEALICVGSSKKRARKA 682
Db   1089 EVARETVEPPEDEPPGSLGSVDENRETLTSLEKESQELSSLGKWNVETRVEDSQQCLQVE 1148

Qy    683 SSSDDEGGPRTL---------GGDSHRAEEASKDKEAGTDAVPAST-------------- 719
Db   1149 EGLQEEQHQESLREVKQELPSSGNQQRWEDVVEGKAVGQEAPLATTGVGTEDKAELHLRG 1208

Qy    720 ---QEQDQAQ-------------GSSSPEPAGSPSEG----EGVSTWESFKRLVTPRKK 758
Db   1209 QGGEEEAAAEGELLQDIVGEAWSLGSSEPKEQRVPAEALDNLEGGA-------LEVPVAQ 1261

Qy    759 SKSKLEEKAED--SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGP 816
Db   1262 SMPEVTERDEDRAQAGEQDSIEVTLGLEAARTGLEL------------EQEVVGLED--P 1307

Qy    817 VEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVS-EELSKTLVHTVSVA 875
Db   1308 RHFAREEAIPPS---LGEESVKAKIAQGLEGPGKEPKEAGALDSGILELPKTSSEALECQ 1364

Qy    876 VIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEET--------- 926
Db   1365 GHEESESMEGWEEEEAS------LETSDHEGSDAPQPRPPETEEDEGAQAALTAPGPKLL 1418

Qy    927 ------PVLTQTLPEGKDAHDDMVTSE--------VDFTSEAVTATETSEALRTEEVTEA 972
Db   1419 EPCSPIPILT-------DAHELQPQAEGIQEAGWQPEAGSEALERVENEPEFGLGEIPEG 1471

Qy    973 -----SGAEETTDMVSAVSQLTDSPDTTEEATP----VQEVESGVLDTEEEERQTQAILQ 1023
Db   1472 LQDWEEGREE--------SEADDLGETLPDSTPLGLYLRSPASPKWDLAGEQR-----LS 1518

Qy   1024 AVADKVKEE--SQVPATQTVQRTGSKALEKVEEVEEDSEVLASEK------EKDVMPKGP 1075
Db   1519 PQGDAGKEDWGPAVPAAQ-----GLSGPPEEEEEQGHGSDLSSEEFEDLGTEASLLPGVP 1573

Qy   1076 VQEAGAEHLAQ-------------GSETGQATPESL--EVPEVTADVDHVAT-------- 1112
Db   1574 KEV--ADHVGQVPPVLQPACWDQGGESDGFADEEESGEEGEEEDADEEGAESGAQWWGSG 1631

Qy   1113 -----CQVIKLQQLMEQAVAPESSETLTDSETNGST------PLADSDTADGTQQDETID 1161
Db   1632 ASGGGCKVQDIAQRGDPVQESVGVSGLWDDGLRGAAANVPALEMVSQDSAEPSGSEES-- 1689

Qy   1162 SQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAA 1221
Db   1690 ---ESASLEGEEGQVTDH--LDAPQEVTSMVPGVGDAFDIGGQSPNLD------------ 1732
```

```
Qy    1222 AVPVLAKTEVG---QEGEVDWLDGEKVKEEQEVF-----------------VHSGPN 1258
Db    1733 SEQVNGKMENGLEQAEGQV-VLDGD---EDQELLLQGQEVGALKVPLVASPVHLGPS 1785
```

RESULT 14
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
;   LENGTH: 3696
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

```
  Query Match             4.5%;  Score 365.5;  DB 4;  Length 3696;
  Best Local Similarity   18.2%;  Pred. No. 1.2e-13;
  Matches  324;  Conservative  309;  Mismatches  694;  Indels  449;  Gaps   75;

Qy      46 DPATKLPQKNGQL----SSVNGVAEQGDVHVQ----EENQEGQEEEVVDEDVGQRESEDV 97
Db    1462 DAITSLSRINAQVVEKESARNAIEQKATQQTQFINNNDNATDEEKEVANNLVIATKQKSL 1521

Qy      98 REKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKV 157
Db    1522 ---DNINSLSSNND-VENAKVAGINEIANVLPATAVKSKAKKDIDQKLAQQINQIQTHQT 1577

Qy     158 FKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETA--------- 208
Db    1578 -----------ATTEEKEAAIQLANQKSNEARTAIQN----EHSNNGVAQAKSNGIHEIE 1622

Qy     209 -VGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEE---------AKDEGE 258
Db    1623 LVMPDAHKKSDAKQSIDNK----YNEQSNT---INTTPDATDEEKQKALDKLKIAKDAGY 1675

Qy     259 EKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFK---KSKEDDLE----- 310
Db    1676 NKVDQAQTNQQVSDAKTEAIDTITNIQ-------ANVAKKPSARVELDSKFEDLKRQINA 1728

Qy     311 --TAEKRKEQEA-----------------EKVDEEEKEKTEPASEEQE-----PAEDTD- 345
Db    1729 TPNATEEEKQDAIQRLNGKRDEVKNLINQDRRDNEVEQHKNIGLQELETIHANPTRKSDA 1788

Qy     346 ----QARLSADYEKVELPLEDQVGDLEASSEEK-------------------CAPLATEV 382
Db    1789 LQELQTKFISQTELIN-------NNKDATNEEKDEAKRLLEISKNKTITNINQAQTNNQV 1841

Qy     383 FDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGV-VVEGTGESLPPEKLAEPQEVPQ 441
Db    1842 DNAKDNGMNEIATIIPATTI-KTDAKTAIDKKAEQQVTIINGNNDATDEEK-AEARKLVE 1899

Qy     442 EA--EPAEELMKS-REMCVSGGDHTQLTDL-----SPEEKTLPKHPEGIVSEVEMLSSQE 493
Db    1900 KAKIEAKSNITNSDTEREVNGAKTNGLEKINNIQPSTQTKTNAK------QEINDKAQEQ 1953

Qy     494 RIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHI----HTESPESADEQK 549
Db    1954 LIQINNTP--------------DATEEEKQEATNRVNAGLAQAIQNINNAHSTQEVNESK 1999

Qy     550 GESSAS----SPEEPEETTCLEKGPLEAPQDGE--AEEGTTSDGEKKREGITPWASFKKM 603
Db    2000 TNSIATIKSVQPNVIKKPTAINSLTQEANNQKTLIGNDGNATDDEKE--------AAKQL 2051

Qy     604 VTPK--KRVRRPSESDKEEELEKVKSATLSS---TDSTVSEMQDEVKTVGE--EQKPEEP 656
Db    2052 VTQKLNEQIQKIHESTQDNQVDNVKAQAITAIKLINANAHKRQDAINILTNLAESKKSDI 2111

Qy     657 KRRVDTSVSWE--ALICVGSSKKRARKASSSDDEGGPRTLGGDSHRA--EEASKDKEAGT 712
Db    2112 RANQDATTEEKNTAIQSIDDTLAQAR-----------NNINGANTNALVDENLEDGKQKL 2160

Qy     713 DAVPASTQEQDQAQGSSSP---EPAGSPSEGEGVST---WESFKRL-------------- 752
Db    2161 QRIVLSTQTKTQAKADIAQAIGQQRSTIDQNQNATTEEKQEALERLNQETNGVNDRIQAA 2220

Qy     753 -----VTPRKKS--------------KSKLEEKAEDSSVEQ---LSTEIEPSREESWVSI 790
Db    2221 LANQNVTDEKNNILETIRNVEPIVIVKPKANEIIRKKAAEQTTLINQNQDATLEEKQIAL 2280

Qy     791 KKFIPGRRK-----KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEA 845
Db    2281 GKLEEVKNEALNQVSQAHSNNDVKIAENNGIAKISEVHPET-----IIKRNA--KQEIEQ 2333

Qy     846 QGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAV------------------TSVEE 888
Db    2334 DAQSQIDTINANNKSTNEEKSAAIDRVNVAKIDAINNITNATTTQLVNDAKNSGNTSISQ 2393

Qy     889 RSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEV 948
Db    2394 ILPS--TAVKTNALAALASEA-------KNKNAIIDQTPNATAEEKEEANNKVDRLQEEA 2444

Qy     949 DFTSEAVTATETSEALRTEEVTEASGAE----ETTDMVSAVSQLTDSPDTTEEATPVQEV 1004
Db    2445 DANILKAHTTDEVNNIKNQAVQNINAVQVEVIKKQNVKNQLNQFIDNQKKIIENTPDATL 2504

Qy    1005 ESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLAS 1064
Db    2505 E----EKAEANRLLQNVLTSTSDEI---ANVDHNNEVDQALDKARPKIEEI-----VPQV 2552

Qy    1065 EKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQ 1124
Db    2553 SKKRDVL--NAIQEAFNSQTQEIQENQEATNEE-------------KTEALNKINQLLNQ 2597

Qy    1125 AVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEE----- 1179
Db    2598 AKV---------------------NIDQAQSNKDVDSAKTRSIQDIEQIQPHPQTKATG 2635

Qy    1180 -----EAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAVPVLAKTEVGQE 1234
Db    2636 RHRLNEKANQQQSTIATHPNSTIEERQEASAKLQEVLK-----------KAIAKIDKGQT 2684

Qy    1235 GEVDWLDGEK-----VKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLS 1289
Db    2685 ND----DVEKTVVNGIAEIENILPATTVKDKAKADVN-----------AEKEQKNLQINS 2729

Qy    1290 LEEGEMETDVEKEKRETKPEQVSEEGEQETA-APEHEGTYGKPVLTLDMPSSERGKALGS 1348
Db    2730 NDEATTE---EKLVASDNLNHVVETTNQAIEDAPD-----------TNQVNVEKNKGIGT 2775

Qy    1349 LGG-SPSLPDQDKA-GCIEVQVQSLDTTVTQTAEAV---------------EKVIETVVI 1391
Db    2776 IRDIQPLVVKKPTAKSKIESAVEKKKTEINQTQNATHDEVREGLNQLNQIHEKAKNDVNQ 2835

Qy    1392 SETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLH 1451
Db    2836 SQTNQQVE-------NAEQNS-------LDQINNFRP--DFSKKRNAVAEIVKAQQNKI- 2878

Qy    1452 PDLQGEISASQRERS------EEE---------DKPDAGPDADGKESTAIEKVLKAEPEIL 1497
Db    2879 DEIEQEFSATQEEKDNALQHLDEQVKEIINSINQANTDNEVDNAKTSGLNNITEYRPEY- 2937

Qy    1498 ELESKSNKIVLNVIQTAVDQFARTETAP---ETHAYDSQTQVPACRLDSR---------- 1544
```

```
Db   2938 ---NKKKNAILKLYDVSDTQEAIINGYPDATEDELQEANSKLNKILLDAKKQIGLAHTNN 2994

Qy   1545 EPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAWAQ 1580

Db   2995 EVDDIYNEVSQ-KMKTILPRVDTKAVARSVLNALAK 3029


RESULT 15
US-08-466-390-4
; Sequence 4, Application US/08466390
; Patent No. 5686562
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/466,390
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  2101 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-466-390-4

  Query Match             4.5%;  Score 361.5;  DB 1;  Length 2101;
  Best Local Similarity   19.6%;  Pred. No. 9.6e-14;
  Matches  364;  Conservative  292;  Mismatches  668;  Indels  529;  Gaps   81;

Qy     53 QKNGQLSSVNGVAEQGDVHVQEENQ--EGQEEEVVDEDVGQRESEDVREKDRV------- 103

Db    357 EKQAQLEKELSAALQDKKCLEEKNEILQGKLSQ-LEEHLSQLQDNPPQEKGEVLGDVLQL 415

Qy    104 -------EEMAANSTAVED-----ITKDGQEET---------------SEIIEQIPASE 135

Db    416 ETLKQEAATLAANNTQLQARVEMLETERGQQEAKLLAERGHFEEEKQQLSSLITDLQSSI 475

Qy    136 NNV----EEMVQPAESQA----------------NDVGFKKVFKFVGFKFTVKKDKNEK 174

Db    476 SNLSQAKEELEQASQAHGARLTAQVASLTSELTTLNATIQQQDQELAGLKQQAKEKQAQL 535

Qy    175 SDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQ 234

Db    536 AQTLQ---------QQEQASQGL-RHQ---VEQLSSLKQKEQQLKEVAEKQEAT--RQD 580

Qy    235 SSTEIPLQAESDQAA--EEEAKDEGEEKQEKEPTKSPE--SPSSPVNSETTSSFKKFFTH 290

Db    581 HAQQLATAAEEREASLRERDAALKQLEALEKEKAAKLEILQQQLQVANEARDSAQTSVTQ 640

Qy    291 GWAGWRKKTSFKKSKED---DLETAEKRKEQEAEKVDEEE---KEKTEPASEEQEPAEDT 344

Db    641 ---AQREKAELSRKVEELQACVETARQEQHEAQAQVAELELQLRSEQQKATEKERVAQEK 697

Qy    345 DQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEV------H 398

Db    698 DQ--LQEQLQALKESLKVTKGSLE---EEK--RRAADALEEQQRCISELKAETRSLVEQH 750

Qy    399 VSTVEKTEEEQGG------------GGEAEGGVVVEGTGESLPPEKLAEPQ-------- 437

Db    751 KRERKELEEERAGRKGLEARLLQLGEAHQAETEVLRRELAEAMAAQHTAESECEQLVKEV 810

Qy    438 -------EVPQEAEP------AEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVS 484

Db    811 AAWRDGYEDSQQEEAQYGAMFQEQLMTLKEECEKARQELQ----EAKEKV-----AGIES 861

Qy    485 EVEMLSSQERIKVQGSPLKKLFS--SSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESP 542

Db    862 HSELQISRQQNK-----LAELHANLARALQQVQEKEVRAQK--LADDLSTLQEKMAATSK 914

Qy    543 ESADEQKGESSASSPEEPEETTCLE--KGPLEA----PQDGEAEEGTTSDGEKKREGITP 596

Db    915 EVA---RLETLVRKAGEQQETASRELVKEPARAGDRQPEWLEEQQG--------RQFCST 963

Qy    597 WASFKKMVTPKKRVRRPSESDKEE---ELEKVKSATLSSTDSTVSE---MQDEVKTVGEE 650

Db    964 QAALQAM-----------EREAEQMGNELERLRAALMESQGQQQEERGQQEREVARLTQE 1012

Qy    651 QKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKE- 709

Db   1013 RGRAQADLALEKAARAELEMRLQNALNEQRVEFATLQEALAHAL------TEKEGKDQEL 1066

Qy    710 ---AGTDAV-----------------PASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESF 749

Db   1067 AKLRGLEAAQIKELEELRQTVKQLKEQLAKKEKEHASGSGAQSEAAGRTEPTGPKL-EAL 1125

Qy    750 KRLVTPRKKSKSKLEEKAE----------------DSSVEQLSTEIEPSREESWVSIKKF 793

Db   1126 RAEVSKLEQQCQKQQEQADSLERSLEAERASRAERDSALETLQGQLEEKAQELGHSQSAL 1185

Qy    794 IPGRRKKRA--DGKQEQATVEDSGPVEI----NEDDPNVPAVVPL-SEYNAVEREKMEAQ 846

Db   1186 ASAQRELAAFRTKVQDHSKAEDEWKAQVARGRQEAERKNSLISSLEEEVSILNRQVLEKE 1245

Qy    847 GNTELPQLLGAVYYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTA 906

Db   1246 G--ESKELKRLVMAESEKSQKLEESCACC-----------RQRQPA-------------- 1278

Qy    907 GEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSE---VDFTSEAVTATETSEA 963

Db   1279 -----TVPELQNAALLCGR-----RCRASGREAEKQRVASENLRQELTSQAERAEELGQE 1328

Qy    964 LRT--EEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAI 1021

Db   1329 LKAWQEKFFQKEQALSTLQL---------------EHTSTQALVSELLPAKHLCQQLQAE 1373

Qy   1022 LQAVADKVKEESQVPATQTVQRTGSKALE----------------KVEEVEEDSEVLASE 1065

Db   1374 QAAAEKRHREELE----QSKQAAGGLRAELLRAQRELGELIPLRQKVAEQERTAQQLRAE 1429

Qy   1066 K-------------------------------------------EKDVMPKGPVQEAGAE 1082

Db   1430 KASYAEQLSMLKKAHGLLAEENRGLGERANLGRQFLEVELDQAREKYVQELAAVRADAET 1489

Qy   1083 HLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGS 1142

Db   1490 RLAEVQREAQSTARELEV--MTAKYEGA------KVKVLEERQRFQEERQKLTAQVEELS 1541

Qy   1143 TPLADSDTADGTQQD--ETIDSQ--DSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPA 1198

Db   1542 KKLADSDQASKVQQQKLKAVQAQGGESQQEAQRFQAQLNELQAQLSQKEQAA-------- 1593

Qy   1199 QEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPN 1258
```

```
Db   1594 -------------EHYKLQMEKAKTHYDAKKQQNQELQEQLRSLEQLQKE---------N 1631

Qy   1259 SQKAADVTYDSEVMGVAGCQEKESTEV-QSLSLEEGEMETDV---EKEKRETKPEQVSEE 1314
            :  |:       :  || : ||: :  : |: :   :|  |    ::: |:     ||:
Db   1632 KELRAEAERLGHELQQAGLKTKEAEQTCRHLTAQVRSLEAQVAHADQQLRDLGKFQVA-- 1689

Qy   1315 GEQETAAPEHEGTYGKPVLTLDMP----SSERGKALG-------------SLGGSPSLPD 1357
               | | :      || | | :      | | | |               |: | |: |
Db   1690 ----TDALKSREPQAKPQLDLSIDSLDLSCEEGTPLSITSKLPRTQPDGTSVPGEPASPI 1745

Qy   1358 QDKAGCIEVQVQSLD----TTVTQTAEA-VEKVIETV--VISETGESPE----------- 1399
            :   :  :|:||:      | :    ::| :|  ::::  |  ::|
Db   1746 SQR---LPPKVESLESLYFTPIPARSQAPLESSLDSLGDVFLDSGRKTRSARRRTTQIIN 1802

Qy   1400 -CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQA--------ESIPIIVTPAPESTL 1450
             :   |   |   ||    :: : |         | ||       :|:  : :|   ::
Db   1803 ITMTKKLDVEEPDSANSSFYSTRSA-------PASQASLRATSSTQSLARLGSPDYGNSA 1855

Qy   1451 HPDLQG--EISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPE-------ILELES 1501
               | |      : |   ||:       | |   |: |   :    : |||        | ||:
Db   1856 LLSLPGYRPTTRSSARRSQAGVSSGAPP---GRNSFYM-GTCQDEPEQLDDWNRIAELQQ 1911

Qy   1502 KSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHP 1561
          : |::     ::|                             |:|| |:     : | :||
Db   1912 R-NRVCPPHLKTCYP------------------------LESR-PSLSLGTITDEEMKTG 1945

Qy   1562 VPQP---REDLQVLTVLEAWA----QPRK---------------------CLPR 1587
            ||     |  :| : : |        | ||                    | ||
Db   1946 DPQETLRRASMQPIQIAEGTGITTRQQRKRVSLEPHQGPGTPESKKATSCFPR 1998
```

Search completed: December 13, 2002, 00:17:10
Job time : 54 secs

```
OM protein - protein search, using sw model

Run on:         December 13, 2002, 00:16:06 ; Search time 224 Seconds
                                              (without alignments)
                                              118.719 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       105981 seqs, 16662342 residues

Total number of hits satisfying chosen parameters:      105981

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       Published_Applications_AA:*
                 1:  /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
                 2:  /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
                 3:  /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
                 4:  /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
                 5:  /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
                 6:  /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
                 7:  /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
                 8:  /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
                 9:  /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
                 10:  /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
                 11:  /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
                 12:  /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
                 13:  /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
                 14:  /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8073	100.0	1596	9	US-09-902-432-4	Sequence 4, Appli
2	5421.5	67.2	1346	9	US-09-902-432-2	Sequence 2, Appli
3	4294	53.2	1781	10	US-09-738-877-3	Sequence 3, Appli
4	462	5.7	1786	9	US-09-742-096-3	Sequence 3, Appli
5	429	5.3	2665	10	US-09-864-761-34248	Sequence 34248, A
6	407.5	5.0	2368	10	US-09-815-242-5635	Sequence 5635, Ap
7	407.5	5.0	2368	10	US-09-815-242-12389	Sequence 12389, A
8	388	4.8	1325	10	US-09-864-761-35612	Sequence 35612, A
9	382	4.7	1879	9	US-09-971-536-70	Sequence 70, Appl
10	377	4.7	2478	10	US-09-815-242-5816	Sequence 5816, Ap
11	377	4.7	2478	10	US-09-815-242-12967	Sequence 12967, A
12	377	4.7	3256	10	US-09-919-172-98	Sequence 98, Appl
13	368	4.6	1400	10	US-09-764-176-7	Sequence 7, Appli
14	354	4.4	1618	9	US-09-963-875-1	Sequence 1, Appli
15	351.5	4.4	617	10	US-09-864-761-36182	Sequence 36182, A
16	339.5	4.2	2139	10	US-09-727-384-6	Sequence 6, Appli
17	334.5	4.1	2344	10	US-09-815-242-12713	Sequence 12713, A
18	323.5	4.0	1884	10	US-09-785-770A-17	Sequence 17, Appl
19	323.5	4.0	1907	10	US-09-785-770A-16	Sequence 16, Appl
20	322	4.0	2310	9	US-09-991-496-120	Sequence 120, App
21	322	4.0	2310	10	US-09-874-923-120	Sequence 120, App
22	314	3.9	1863	9	US-09-734-672-4	Sequence 4, Appli
23	313.5	3.9	2843	9	US-09-987-482-1	Sequence 1, Appli
24	312.5	3.9	2843	8	US-08-681-219-32	Sequence 32, Appl
25	308.5	3.8	26926	9	US-09-759-508B-2	Sequence 2, Appli
26	306	3.8	1863	9	US-09-734-672-2	Sequence 2, Appli
27	306	3.8	1863	9	US-09-734-672-6	Sequence 6, Appli
28	305	3.8	3158	10	US-09-815-242-12611	Sequence 12611, A
29	304	3.8	6281	10	US-09-815-242-12996	Sequence 12996, A
30	296.5	3.7	1285	10	US-09-982-091A-2	Sequence 2, Appli
31	292.5	3.6	350	9	US-10-101-487-58	Sequence 58, Appl
32	290.5	3.6	1597	9	US-09-832-292-35	Sequence 35, Appl
33	285.5	3.5	2437	10	US-09-815-242-5834	Sequence 5834, Ap
34	284	3.5	1501	10	US-09-924-154-17	Sequence 17, Appl
35	281.5	3.5	1065	10	US-09-864-761-43248	Sequence 43248, A
36	279.5	3.5	2828	10	US-09-905-129-21	Sequence 21, Appl
37	279.5	3.5	2828	10	US-09-991-630-21	Sequence 21, Appl
38	278	3.4	2076	10	US-09-815-242-5815	Sequence 5815, Ap
39	278	3.4	2186	10	US-09-815-242-12913	Sequence 12913, A
40	277.5	3.4	2025	10	US-09-815-242-5703	Sequence 5703, Ap
41	274	3.4	868	9	US-09-884-001-19	Sequence 19, Appl
42	267.5	3.3	1332	10	US-09-982-091A-4	Sequence 4, Appli
43	266.5	3.3	1192	10	US-09-758-140-6	Sequence 6, Appli
44	266.5	3.3	1192	10	US-09-972-599A-6	Sequence 6, Appli
45	265.5	3.3	1192	10	US-09-789-386-2	Sequence 2, Appli

ALIGNMENTS

```
RESULT 1
US-09-902-432-4
; Sequence 4, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 1596
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-09-902-432-4

  Query Match             100.0%;  Score 8073;  DB 9;  Length 1596;
  Best Local Similarity   100.0%;  Pred. No. 0;
  Matches 1596;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60

Qy       61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120

Qy      121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy      181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
```

```
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240

Qy     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300

Qy     301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360

Qy     361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420

Qy     421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480

Qy     481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540

Qy     541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600

Qy     601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Qy     661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720

Qy     721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780

Qy     781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840

Qy     841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Qy     901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Qy     961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Qy    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080

Qy    1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140

Qy    1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200

Qy    1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260

Qy    1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320

Qy    1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380

Qy    1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440

Qy    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500

Qy    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560

Qy    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
            ||||||||||||||||||||||||||||||||||||
Db    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
```

RESULT 2
US-09-902-432-2
; Sequence 2, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
; APPLICANT: Irwin H. Gelman
; APPLICANT: Susan G. Jaken
; TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
; FILE REFERENCE: A30558-A-FWC-A 070156.0597
; CURRENT APPLICATION NUMBER: US/09/902,432
; CURRENT FILING DATE: 2002-04-08
; PRIOR APPLICATION NUMBER: 08/978,277
; PRIOR FILING DATE: 1997-11-25
; PRIOR APPLICATION NUMBER: 08/665,401
; PRIOR FILING DATE: 1996-06-18
; PRIOR APPLICATION NUMBER: 08/635,121
; PRIOR FILING DATE: 1996-04-19
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 1346
;   TYPE: PRT
;   ORGANISM: Rattus norvegicus
US-09-902-432-2

```
  Query Match             67.2%;  Score 5421.5;  DB 9;  Length 1346;
  Best Local Similarity   91.6%;  Pred. No. 6.4e-229;
  Matches 1098;  Conservative   24;  Mismatches   62;  Indels   15;  Gaps    7;

Qy     387 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 446
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 60

Qy     447 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 506
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 120

Qy     507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 180

Qy     567 EKGPLEAPQDGEAEEGTTSDGEKKRE----GITPWASFKKMVTPKKRVRRPSESDKEEEL 622
           |||||||: |:  :       ||||:     ||        :|     :: ||||||||||||
Db     181 EKGPLEAPRMGKLRKELLR-GEKKRKDHSLGI-----LQKDGDTQETVRRPSESDKEEEL 234

Qy     623 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 682
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     235 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 294

Qy     683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 742
```

```
Db     295 SSSDIR-GPRTLGGGQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 353

Qy     743 VSTWESFKRLVTPRKKSKSKLEEK-AEDSSVEQLSTEIEPSREESWVS--IKKFIPGRRK 799

Db     354 VSTWESFKRLVTPRKKSKSKLEEKEAGRTLVVGAGCPLRSNRVEKNLGFPLRNSSPDGGR 413

Qy     800 KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVY 859

Db     414 KGQMGRQEQATVEDSGPVEINEDEPDVPAVVPLSEYDAVEREKMEAQGNAELPSCWGCV- 472

Qy     860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919

Db     473 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 532

Qy     920 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 979

Db     533 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 592

Qy     980 DMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039

Db     593 DMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 652

Qy    1040 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 1099

Db     653 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 712

Qy    1100 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 1159

Db     713 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 772

Qy    1160 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELT 1219

Db     773 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELA 832

Qy    1220 AAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279

Db     833 AAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 892

Qy    1280 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPS 1339

Db     893 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVLTLDMPS 952

Qy    1340 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399

Db     953 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1012

Qy    1400 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1459

Db    1013 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1072

Qy    1460 ASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1519

Db    1073 ASQRERSEEEDKPDAGPDADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1132

Qy    1520 RTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAW 1578

Db    1133 RTETAPETHAYDSQTQVPAMRLDSREPNRCWTKMKVAKMKHPVPQPREDLQVLTVLEAW 1191
```

RESULT 3
US-09-738-877-3
; Sequence 3, Application US/09738877
; Patent No. US20020015970A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan
; APPLICANT: Weiss, Stephen J.
; APPLICANT: Glynne, Richard
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET
; TITLE OF INVENTION: SCREENING FOR ANGIOGENESIS MODULATORS
; FILE REFERENCE: A-69806/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/738,877
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,425
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT/US 00/22061
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1781
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-738-877-3

```
  Query Match             53.2%;  Score 4294;  DB 10;  Length 1781;
  Best Local Similarity   57.9%;  Pred. No. 1e-179;
  Matches  959;  Conservative  187;  Mismatches  401;  Indels  108;  Gaps   35;

Qy       1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56

Db       1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy      57 QLSSVNGVAEQGDVHVQEENQEGQ-------------EEEVVDEDVGQRESEDVREKDR 102

Db      61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy     103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162

Db     121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy     163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222

Db     180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy     223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282

Db     235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy     283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337

Db     294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy     338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392

Db     354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy     393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452

Db     414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy     453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512

Db     466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy     513 KLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572

Db     526 KLSGKKQKGKR-GGDDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy     573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632

Db     585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy     633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692

Db     645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRRSSSDEEGGPK 704

Qy     693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752

Db     705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy     753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
```

```
Db     765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy     809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db     825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy     866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db     885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy     925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db     945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy     985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db    1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy    1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db    1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy    1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db    1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy    1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db    1179 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1237

Qy    1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db    1238 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1295

Qy    1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db    1296 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1355

Qy    1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db    1356 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1415

Qy    1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db    1416 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1475

Qy    1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db    1476 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1532

Qy    1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547
Db    1533 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1590

Qy    1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db    1591 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1625
```

RESULT 4
US-09-742-096-3
; Sequence 3, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773US0DIV
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1786
; TYPE: PRT
; ORGANISM: P. falciparum
US-09-742-096-3

```
  Query Match             5.7%;  Score 462;  DB 9;  Length 1786;
  Best Local Similarity   19.2%;  Pred. No. 5.6e-13;
  Matches 325;  Conservative 291;  Mismatches 644;  Indels 436;  Gaps 69;

Qy      53 QKNGQLSSVNGVAEQGDVHVQEENQEGQE-EEVVDEDVGQRESEDVREK---DRVEE--- 105
Db      95 QVNGELAS-----EEVKEKILDLLEEGNTLTESVDDNKNLEEAEDIKENILLSNIEEPKE 149

Qy     106 ---------MAANSTAVEDITKDGQ---EETSEIIEQIPASENNVEEMVQPAESQANDVG 153
Db     150 NIIDNLLNNIGQNSEKQESVSENVQVSDELFNELLNSVDVNGEVKENILE--ESQVNDDI 207

Qy     154 FKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASV--GAGDHQEPSVETAVGE 211
Db     208 FNSLVK------SVQQEQQHNVEEKVEESVEENDEESVEENVEENVEENDDGSVASSVEE 261

Qy     212 S--ASKESELKQSTEKQEGTLKQE------------------QSSTEIPLQAESDQAAEE 251
Db     262 SIASSVDESIDSSIEENVAPTVEEIVAPSVVESVAPSVEESVEENVEESVAENVEESVAE 321

Qy     252 EAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLE- 310
Db     322 NVEESVAENVEESVAENVEEIVAPTVEEIVAPTVEEIVAPSVVESVAPSVEESVEENVEE 381

Qy     311 -TAEKRKEQEAEKVDEEEKEKTEPA---SEEQEPAEDTDQARLSADYEKVELPLEDQVGD 366
Db     382 SVAENVEESVAENVEESVAENVEESVAENVEESVAENVEEI-VAPTVEEIVAPTVEEI-- 438

Qy     367 LEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGE 426
Db     439 VAPSVVESVAPSVEESVEENVE--ESVAENVEESVAENVEES----------VAENVEE 485

Qy     427 SLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEV 486
Db     486 SV--------------AENVEESVAE----------------NVEESVAENVEESVAENV 515

Qy     487 EMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESAD 546
Db     516 EEIVAPTVEEIVAPTVEEIVAPSVVESVA---------------PSVEESVEENVEESVA 560

Qy     547 EQKGESSASSPEEP-----EETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFK 601
Db     561 ENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAPTV------EEIVAP--SVV 612

Qy     602 KMVTPKKRVRRPSESDKEEEL-EKVKSATLSSTDSTVSEMQDEV--KTVGEEQKPEEPKR 658
Db     613 ESVAPS--VEESVEENVEESVAENVEESVAENVEESVAENVEEIVAPTVEEIVAP----- 665

Qy     659 RVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEE--ASKDKEAGTDAVP 716
Db     666 ------TVEEIVAPSVVESVAPSVEESVEE-----------NVEESVAENVEESVAENVE 708

Qy     717 ASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTP------RKKSKSKLEEKAEDS 770
Db     709 ESVAENVEESVAENVEEIVAPTVEEIVA--PTVEEIVAPSVVESVAPSVEESVEENVEES 766

Qy     771 SVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQA--TVEDSGPVEINEDDPNVPA 828
Db     767 VAENVEESVAENVEES--------------VAENVEESVAPTVEEIVAPSVEE------S 806

Qy     829 VVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEE 888
```

```
Db    807 VAPSVEESVAENVATNLSDNL-LSNLLGGI-ETEEIKDSILNEIE-----------EVKE 853

Qy    889 RSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEG-KDAHDDMVTSE 947

Db    854 N----VVTTILENVEETTAESVTTFSNILEE---IQENTITNDTIEEKLEELHENVLSAA 906

Qy    948 VDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQ--------LTDSPDTTEEAT 999

Db    907 LENTQSEEEKKEVIDVI--EEVKE----EVATTLIETVEQAEEKSANTITEIFENLEE-- 958

Qy   1000 PVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKA-----LEKVEE 1054

Db    959 --NAVESNENVAENLEKLNETVFNTVLDKVEETVEISGESLENNEMDKAFFSEIFDNVKG 1016

Qy   1055 VEED-----------SEVLASEKEKDV--------------MPKG------------PVQ 1077

Db   1017 IQENLLTGMFRSIETSIVIQSEEKVDLNENVVSSILDNIENMKEGLLNKLENISSTEGVQ 1076

Qy   1078 EAGAEHLAQG------------------SETGQATPESLEVPEVTADVDHVATCQVIK-- 1117

Db   1077 ETVTEHVEQNVYVDVDVPAMKDQFLGILNEAGGLKEMFFNLEDVFKSESDVITVEEIKDE 1136

Qy   1118 -LQQLMEQ---AVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQ 1173

Db   1137 PVQKEVEKETVSIIEEMEENIVDVLEEEKEDLTDK-MIDAVEESIEISSDSKEETESIKD 1195

Qy   1174 SQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLE--PTQQELTAAAVPV------ 1225

Db   1196 K---EKDVSLVVEE----------VQDNDMDESVEKVLELKNMEEELMKDAVEINDITSK 1242

Qy   1226 -LAKTEVGQEGEVDWL-DGEKVKE---------EQEVFVHSGPNSQKAADVTYD-----SE 1270

Db   1243 LIEETQELNEVEADLIKDMEKLKELEKALSEDSKEIIDAKDDTLEKVIEEEHDITTTLDE 1302

Qy   1271 VMGVAGCQE--------------------KESTEVQSLSLEEGE----METDVEKEKRET 1306

Db   1303 VVELKDVEEDKIEKVSDLKDLEEDILKEVKEIKELESEILEDYKELKTIETDILEEKKEI 1362

Qy   1307 KPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSE------------------------- 1341

Db   1363 EKDHFEKFEEEAEEIKDLEADILKEVSSLEVEEEKKLEEVHELKEEVEHIISGDAHIKGL 1422

Qy   1342 -----------RGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVV 1390

Db   1423 EEDDLEEVDDLKGSILDMLKGDMELGDMDK---------ESLEDVTTKLGERVES-LKDVL 1473

Qy   1391 ISETGESPECVGAH-----------LL-------PAEKSSATGGHWTLQHAE--DTVPLG 1430

Db   1474 SSALGMDEEQMKTRKKAQRPKLEEVLLKEEVKEEPKKKITKKKVRFDIKDKEPKDEI-VE 1532

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490

Db   1533 VEMKDEDI--------EEDVEEDIEEDIEEDKVEDIDEDIDEDIGEDKD----EVIDLIV 1580

Qy   1491 KAEPEILELESKSNKI 1506

Db   1581 QKEKRIEKVKAKKKKL 1596
```

RESULT 5
US-09-864-761-34248
; Sequence 34248, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 34248
;   LENGTH: 2665
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL034555.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
;   OTHER INFORMATION: EST_HUMAN HIT: AU117052.1, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUE 3.00e-10
US-09-864-761-34248

```
  Query Match             5.3%;  Score 429;  DB 10;  Length 2665;
  Best Local Similarity   19.6%;  Pred. No. 2.4e-11;
  Matches  372;  Conservative  270;  Mismatches  755;  Indels  504;  Gaps   83;

Qy      9 QRSPEQPAGSD----------TPSELVLSGHGPAAEASG---------AAGDPADADPAT- 49

Db    301 QVSQTEPAKSDLSKLESVRMKVPKEKGLSSHVEVVEKEGRLKARKHLKPEQPADGVSAVD 360

Qy     50 --KLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMA 107

Db    361 LEKLEARKRRFADSNLKAEKQKPEVKKSSPEMEDARVL-----SKKQPDVSSREVI---- 411

Qy    108 ANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTV 167

Db    412 --------LLREGEAERKPVRKEILKRES-------------------KKI---------- 435
```

```
Qy    168 KKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQE 227
Db    436 ------KLDRLNTVASPKDCQELASISVGSGSRPSSDLQARLGELAG------ESVENQE 483

Qy    228 GTLKQEQSSTEIP-------LQAESDQAAEEE--AKDEGEEKQEKEPTKSPESPSSPVNS 278
Db    484 -----VQSKKPIPSKPQLKQLQVLDDQGPEREDVRKNYCSLRDETPERKSGQEKSHSVNT 538

Qy    279 ET--------TSSFKKFFTHGWAGWRKKTSFKKSK-----EDDLETAEKR---------- 315
Db    539 EEKIGIDIDHTQSYRKQMEQSRRKQQMEMEIAKSEKFGSPKKDVDEYERRSLVHEVGKPP 598

Qy    316 ----------KEQEAEKVD-------EEEKEKTEPASEEQEPAEDTDQARLSADYEKVEL 358
Db    599 QDVTDDSPPSKKKRMDHVDFDICTKRERNYRSSRQISEDSERTGGSPSVRHGSFHED--- 655

Qy    359 PLEDQVGDLEASS-------EEKCAP----------LATEVFDEKMEAHQEVVAEVHVST 401
Db    656 --EDPIGSPRLLSVKGSPKVDEKVLPYSNITVREESLKFNPYDSSRREQMADMAKIKLSV 713

Qy    402 VEKTEEEQGGGGEAE---GGVVVEGTGESLPPEKL-------AEPQEVPQEAEPAEE--- 448
Db    714 LNSEDELNRWDSQMKQDAGRFDVSFPNSIIKRDSLRKRSVRDLEPGEVPSDSDEDGEHKS 773

Qy    449 ----------------LMKSREMCVSGGDHTQLTDLSPE-------EKTLPKHPEGIVSE 485
Db    774 HSPRASALYESSRLSFLLRDREDKLRERDERLSSSLERNKFYSFALDKTITPDTKALLER 833

Qy    486 VEMLSS---------------------QERIKVQGSP--------LKKLFSSSGLKKLSG 516
Db    834 AKSLSSSREENWSFLDWDSRFANFRNNKDKEKVDSAPRPIPSWYMKKKKIRTDSEGKMDD 893

Qy    517 KKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQD 576
Db    894 KKEDHK-----EEEQERQELFASRFLHSSIFEQDSKRLQHLERKEEDSDFISGRIYGKQT 948

Qy    577 GEAEEGTTSDGEKKREGITPWAS-FKKMV----TPKKRVRRPSESDKEEELEKVKSATLS 631
Db    949 SEGANSTT---DSIQEPVVLFHSRFMELTRMQQKEKEKDQKPKEVEKQEDTENHPKTPES 1005

Qy    632 STDSTVSEMQ-------DEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASS 684
Db   1006 APENKDSELKTPPSVGPPSVTVVTLESAPSALEKTTGDKTVEAPLV---TEEKTVEPATV 1062

Qy    685 SDD-----EGGPRTLGGDSHRAEE------ASKDKEAGTDAVPASTQEQDQAQGSSSPEP 733
Db   1063 SEEAKPASEPAPAPV----EQLEQVDLPPGADPDKEAA--MMPAGVEE-----GSSGDQP 1111

Qy    734 ----AGSPSEGEGVSTWE--------SFKRLVTPRKKSKSKLEEKAE--DSSVE------ 773
Db   1112 PYLDAKPPTPGASFSQAESNVDPEPDSTQPLSKPAQKSEEANEPKAEKPDATADAEPDAN 1171

Qy    774 ---QLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVV 830
Db   1172 QKAEAAPESQPPASED-LEVDPPVAAKDKKPNKSKRSKTPVQ-AAAVSIVE-----KPVT 1224

Qy    831 PLSEYNAVEREKMEAQGNT--ELPQLLGAVYVSEELSKT--------LVHTVSVAVIDGT 880
Db   1225 RKSE--RIDREKLKRSNSPRGEAQKLLELKMEAEKITRTASKNSAADLEHPEPSLPLSRT 1282

Qy    881 RA--VTSV-------EERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAE--ETPVL 929
Db   1283 RRRNVRSVYATMGDHENRSP------VKEPVEQ---------PRVTRKRLERELQEAAAV 1327

Qy    930 TQT-----LPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAE----ETTD 980
Db   1328 PTTPRRGRPPKTRRRADEEEENEAKEPAETLKPPEGWRSPRSQKTAAGGGPQGKKGKNEP 1387

Qy    981 MVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQT 1040
Db   1388 KVDAT-----RPEATTEVGPQIGVKESSMEPKAAEEEAGSEQKRDRKDAGTDKNPPETAP 1442

Qy   1041 VQRTGSKALEKVEEVEEDS-------------EVLASEKEKD--VMPKGPVQEAGAEHL 1084
Db   1443 VE-----VVEKKPAPEKNSKSKRGRSRNSRLAVDKSASLKNVDAAVSPRGAAAQAGER-- 1495

Qy   1085 AQGSETG--QATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGS 1142
Db   1496 ----ESGVVAVSPEKSESPQKEDGLS----------SQLKSDPVDPDKEPEKEDVSASGP 1541

Qy   1143 TPLADSDTADGTQQDETIDS----QDSKATAAVRQS--------------QVTEEEAAT 1183
Db   1542 SPEA-TQLAKQMELEQAVEHIAKLAEASASAAYKADAPEGLAPEDRDKPAHQASETELAA 1600

Qy   1184 AQK---EEPSTLPNNVPAQEEH-GEE-------PGRDVLEPTQQEL-TAAAVPVLAKTEV 1231
Db   1601 AIGSIINDISGEPENFPAPPPYPGESQTDLQPPAGAQALQPSEEGMETDEAVSGILETEA 1660

Qy   1232 GQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLE 1291
Db   1661 ATESSRPPVNAPDPS--------AGPTDTKEARGNSSETSHSVPEAKGSKEVEVTLVRKD 1712

Qy   1292 EGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGG 1351
Db   1713 KGRQKTTRSRRKRNTNKKVVA---PVESHVPESNQAQG------ESPAANEGTTV----Q 1759

Qy   1352 SPSLPDQDKAGCIEVQVQSLDTTVTQTA-------EAVEKVIETVVISETGESPECVGAH 1404
Db   1760 HPEAPQEEK------QSEKPHSTPPQSCTSDLSKIPSTENSSQEISVEERTPTKASVPPD 1813

Qy   1405 L----LPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPA--------------- 1445
Db   1814 LPPPPQPAPVDEEPQARFRVHSIIESDPVTPPSD-PSIPIPTLPSVTAAKLSPPVASGGI 1872

Qy   1446 PESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKE----STAIEKVLKAEPEILELES 1501
Db   1873 PHQSPPTKVTEWITRQEEPRAQSTPSPALPPDTKASDVDTSSSTLRKIL-MDPKYVSATS 1931

Qy   1502 KSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHP 1561
Db   1932 VTSTSVTTAIAEPV------SAAPCLH------EAPPPPVDSKKP-------LEEKTAPP 1972

Qy   1562 VPQPREDLQVLTVLEAWAQPRKCLPRLQLK-------APVS 1595
Db   1973 VTN-NSEIQASEVLVA-ADKEKVAPVIAPKITSVISRMPVS 2011
```

RESULT 6
US-09-815-242-5635
; Sequence 5635, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27

```
;  PRIOR APPLICATION NUMBER: 60/257,931
;  PRIOR FILING DATE: 2000-12-22
;  PRIOR APPLICATION NUMBER: 60/269,308
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5635
;   LENGTH: 2368
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-5635

  Query Match             5.0%;  Score 407.5;  DB 10;  Length 2368;
  Best Local Similarity   17.5%;  Pred. No. 1.8e-10;
  Matches  317;  Conservative  293;  Mismatches  663;  Indels  541;  Gaps   71;

Qy     47 PATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEE-EVVDEDVGQRESEDVREKD---- 101
Db    677 PTVKPQAKQDIIQAVTTRKQQ----IKKSNASLQDEKDVANDKIGKIETKAIKDIDAATT 732

Qy    102 RVEEMAANSTAVEDITKDGQ---------EETSEI----IEQIPASENNVEEMVQPAESQ 148
Db    733 NAQVEAIKTKAINDINQTAPATTAKAAALEEFDEVVQAQIDQAPLNPDTTNEEVAEAIER 792

Qy    149 AN--DVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLT----VKKDEGEGAEASVGAGDHQE 202
Db    793 INAAKVSGVKAIEATTAQDLERVKNEEISKIENITDSTQTKMDAYNEVKQAATARKTQN 852

Qy    203 PSVETAVGESASKESELKQSTEK------QEGTLKQEQSSTE---------IPLQAESDQ 247
Db    853 ATVSNATNEEVAEADAAVEAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIQTQAKVKP 912

Qy    248 AAEEEAKDE-GEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSK- 305
Db    913 AADTEVENAYNTRKQEIQNS-----------NASTTEEKQAAYT--------ELDTKKQEA 954

Qy    306 ---------EDDLETAE---------------KRKEQEAEKVDEEEKEKT------EPAS 335
Db    955 RTNLDAANTNSDVTTAKDNGIAAINQVQAATTKKSDAKAEIAQKASERKTAIEAMNDSTT 1014

Qy    336 EEQEPAED-TDQARLSADYEKVELPLEDQVGDLEASSEEKCA---------PLATEVFDE 385
Db   1015 EEQQAAKDKVDQAVVTANADIDNAAANTDVDNAKTTNEATIAAITPDANVKPTAKQAIAD 1074

Qy    386 KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEP 445
Db   1075 KVQAQETAI-------------------DANNGATTE--------EKAAAKQQVQTEKTT 1107

Qy    446 AEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKL 505
Db   1108 ADTAI--------DGAHTN-------------------AEVEAAKNAEIAKIEAIQPATT 1140

Qy    506 FSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTE------------------------S 541
Db   1141 TKDNAKQAIATKANERKTA------IAQTQDITAEEIAAANANVDNAVTQANNNIEAANS 1194

Qy    542 PESADEQK--GESSAS--SPEEPEETTCL--EKGPLEAPQDGEAEEGTTSDGEKKREGIT 595
Db   1195 QNDVDQAKTTGEASIDQVTPTVNKKATAVTDAKNNITAATDDNGVDTAKDAGKNSIQSTQ 1254

Qy    596 PWASFKK---------MVTPKKRVRRPSESDKEE------------------------- 620
Db   1255 PATAVKSNAKNDVDQAVTTQNQAIDNTTGATTEEKNAAKDLVLKAKEKAYQDILNAQTTN 1314

Qy    621 ELEKVKSATLS-----STDSTVSEM-QDEVKTVGEEQK-------------PEEPKRRVD 661
Db   1315 DVTQIKDQAVADVQGITADTTIKDVAKDELATKAREQKALIAQTADATTEEKEQANQQVD 1374

Qy    662 TSVSW--------EALICVGSSKKRARKA-----SSSDDEGGPRT-------------LG 695
Db   1375 AELTQGNQNIENAQSIDDVNTAKDNAIQAIDPIQASTDVKTNARAELLTEMQNKITEILN 1434

Qy    696 GDSHRAEEASKD-------KEAGTDAVPAST-------------QEQDQAQGSSSPEPAG 735
Db   1435 NNETTNEEKGNDIGPVRAAYEEGLNNINAATTTGDVTTAKDTAVQKVQQLHANPVKKPAG 1494

Qy    736 SPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAE-DSSVEQLSTEIEPSREESWV------ 788
Db   1495 KTALDQAAAD-KKTQIEQTPNASQQEINDAKQEVDTELNQAKTNIDQSSTDEYVDNAVKE 1553

Qy    789 ------SIKKFIPGRR--------------------------------------KKRADG 804
Db   1554 GKAKINAVKTFSEYKKDALAKIEAAYNAKVTEADNSNASTSSEIAEAKQKLAELKQTADQ 1613

Qy    805 KQEQATVEDSGPVEINEDDPNV-----------PAVVPLSEYNAVEREKMEAQGNTELPQ 853
Db   1614 NVNQATSKDDIEVQIHNDLDNINDYTIPTGKKESATTDLYAYADQKKNNISADTNATQDE 1673

Qy    854 LLGAVYVSEELSKTLVHTVS---------------VAVIDGTRAVTSVEERSPSWISASV 898
Db   1674 KQQAIKQVDQNVQTALESINNGVDNGDVDDALTQGKAAIDAIQVDATVKPKANQVIDAKA 1733

Qy    899 TEPLE----------HTAGEAMPPVEEVTE--KDIIAEETPV--LTQTLPEGKDAHDDMV 944
Db   1734 EETKESIDQSDQLTAEEKTEALAMIKQITDQAKQGITDATTAEVEKAKAQGLEAFDNI- 1792

Qy    945 TSEVDFTSEAVTATETSEALRTEEVTEASGAEETTD------------MVSAVSQLTDSP 992
Db   1793 --QIDSTEKQKAIEELETALDQIEAGVNVDADATTEEKEAFTNALEDILSKATEDISDQT 1850

Qy    993 DTTEEATPVQEVESGVLDTEEEERQTQAI----LQAVADKVKEESQV-------PATQTV 1041
Db   1851 TNAEIAT----VKNSALEQLKAQRINPVVKKNALEAIREVVNKQIEIIKNADADASAKEI 1906

Qy   1042 QRTG-----SKALEKVEEVEEDSEVLASEKEKDVMPK----GPVQEAGAEHLAQGSE--- 1089
Db   1907 ARTDLGRYFDRFADKLDKTQTNTEV--AELQNVTIPAIEAIVPQNDPNANDTNSGSDNND 1964

Qy   1090 -----TGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTP 1144
Db   1965 ATANSNANATPENTGQPNVTESTDNANA----------------DTSSTTTNNQNDAATG 2008

Qy   1145 LADSDTADGTQQDETID--SQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNN----VPA 1198
Db   2009 ETTATSANSSATDDANDKPQANNNSSADTSTNSPTMDNDVTSKPEVEST--NNGTTDKPA 2066

Qy   1199 QEEHGEEPGRDVLE------------PTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK 1246
Db   2067 TEADNATPAESATNNNSTTTATNENAPTESTATA---PTTAST--GAESSADSKDNASVN 2121

Qy   1247 E-EQEVFVHSG------------PNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEG 1293
Db   2122 DSKQNAEVNNSAESQSTNGKVAQPKSENKAKAEKDGRDSTNQSMVESTTETLPSADITEP 2181

Qy   1294 EMETDVEKEKRETKPEQVSEE---GEQETAAPEHEGTYG---------KPVLTLDMPSSER 1342
Db   2182 KVSSNTSKDKEESTTSQTDAEQHNSDTNVASNEADKSEGNVDTDVSNKP--STSKPSEAK 2239

Qy   1343 GKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVG 1402
Db   2240 DKA-------TSTEDSQKADMATADTKDNQASIGATADVNNKATQN------------DG 2280

Qy   1403 AHLLPAEKSSATGGHWTLQHAEDTVPLGPES-QAESIPIIVTPAPESTLHPDLQGEISAS 1461
Db   2281 ANASPA-----------------TVSKGSNSANQDMLNVTKTKENKANAKSAQQGKVN-- 2321

Qy   1462 QRERSEEEDKPDAG 1475
Db   2322 -KPKQQAKTLPDTG 2334

RESULT 7
US-09-815-242-12389
; Sequence 12389, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Haselbeck, Robert
; APPLICANT:  Ohlsen, Kari L.
; APPLICANT:  Zyskind, Judith W.
; APPLICANT:  Wall, Daniel
; APPLICANT:  Trawick, John D.
; APPLICANT:  Carr, Grant J.
; APPLICANT:  Yamamoto, Robert T.
; APPLICANT:  Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION:  Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE:  2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12389
;   LENGTH: 2368
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12389

  Query Match             5.0%;  Score 407.5;  DB 10;  Length 2368;
  Best Local Similarity   17.5%;  Pred. No. 1.8e-10;
  Matches  317;  Conservative  293;  Mismatches  663;  Indels  541;  Gaps   71;

Qy       47 PATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEE-EVVDEDVGQRESEDVREKD---- 101
            |  |   |    : :|    :|    ::: |   |:| :| :: :|: |:: ::: |
Db      677 PTVKPQAKQDIIQAVTTRKQQ----IKKSNASLQDEKDVANDKIGKIETKAIKDIDAATT 732

Qy      102 RVEEMAANSTAVEDITKDGQ---------EETSEI----IEQIPASENNVEEMVQPAESQ 148
             :  |  : |: || :             ||  |:    |:| |  : :  |  |  :
Db      733 NAQVEAIKTKAINDINQTAPATTAKAAALEEFDEVVQAQIDQAPLNPDTTNEEVAEAIER 792

Qy      149 AN--DVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLT----VKKDEGEGAEASVGAGDHQE 202
             |   |  |   :  ::: |||:   :: :|   |  |        : :  |   |
Db      793 INAAKVSGVKAIEATTAQDLERVKNEEISKIENITDSTQTKMDAYNEVKQAATARKTQN 852

Qy      203 PSVETAVGESASKESELKQSTEK------QEGTLKQEQSSTE---------IPLQAESDQ 247
             :|  |  |  ::     :: :|      |      |||  : |:         |   ||:
Db      853 ATVSNATNEEVAEADAAVEAAQKQGLHDIQVVKSKQEVADTKSKVLDKINAIQTQAKVKP 912

Qy      248 AAEEEAKDE-GEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSK- 305
            ||: |  ::     ||| : :         |: ||    : :|          :   || :
Db      913 AADTEVENAYNTRKQEIQNS----------NASTTEEKQAAYT--------ELDTKKQEA 954

Qy      306 ---------EDDLETAE---------------KRKEQEAEKVDEEEKEKT------EPAS 335
                     |: ||:                |: : :||    :  :  ||      :   :
Db      955 RTNLDAANTNSDVTTAKDNGIAAINQVQAATTKKSDAKAEIAQKASERKTAIEAMNDSTT 1014

Qy      336 EEQEPAED-TDQARLSADYEKVELPLEDQVGDLEASSEEKCA---------PLATEVFDE 385
            |||: |:|  |||  ::|:|: :        | : : :|             | |  : :
Db     1015 EEQQAAKDKVDQAVVTANADIDNAAANTDVDNAKTTNEATIAAITPDANVKPTAKQAIAD 1074

Qy      386 KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEP 445
            |::|  :   :                   :|  |    |        || |  |:|  |
Db     1075 KVQAQETAI-------------------DANNGATTE--------EKAAAKQQVQTEKTT 1107

Qy      446 AEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKL 505
            |:  :           | ||                           :|||   :  |  |::
Db     1108 ADTAI--------DGAHTN------------------AEVEAAKNAEIAKIEAIQPATT 1140

Qy      506 FSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE-----------------------S 541
              :  : :: |  :  |             : |  |   |                   |
Db     1141 TKDNAKQAIATKANERKTA------IAQTQDITAEEIAAANANVDNAVTQANNNIEAANS 1194

Qy      542 PESADEQK--GESSAS--SPEEPEETTCL--EKGPLEAPQDGEAEEGTTSDGEKKREGIT 595
                 |:  |  ||:|     :|    :: |  :     |  :  |   |      :     |:   :
Db     1195 QNDVDQAKTTGEASIDQVTPTVNKKATAVTDAKNNITAATDDNGVDTAKDAGKNSIQSTQ 1254

Qy      596 PWASFKK---------MVTPKKRVRRPSESDKEE------------------------- 620
            |  :  |            :  |  :  :    : :  ||
Db     1255 PATAVKSNAKNDVDQAVTTQNQAIDNTTGATTEEKNAAKDLVLKAKEKAYQDILNAQTTN 1314

Qy      621 ELEKVKSATLS-----STDSTVSEM-QDEVKTVGEEQK------------PEEPKRRVD 661
            :: ::|    ::         : |:|: :: :||: |     |||              |:   ::||
Db     1315 DVTQIKDQAVADVQGITADTTIKDVAKDELATKAREQKALIAQTADATTEEKEQANQQVD 1374

Qy      662 TSVSW--------EALICVGSSKKRARKA-----SSSDDEGGPRT------------LG 695
             ::          :::   |  ::|  |  |:        :|:|  :       |
Db     1375 AELTQGNQNIENAQSIDDVNTAKDNAIQAIDPIQASTDVKTNARAELLTEMQNKITEILN 1434

Qy      696 GDSHRAEEASKD-------KEAGTDAVPAST------------QEQDQAQGSSSPEPAG 735
             :      ||   |          | |  : :  |:|              |:   |     :  :|||
Db     1435 NNETTNEEKGNDIGPVRAAYEEGLNNINAATTTGDVTTAKDTAVQKVQQLHANPVKKPAG 1494

Qy      736 SPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAE-DSSVEQLSTEIEPSREESWV------ 788
              :   :   :   :    :    ||      :      :  | |  |: :  |   | |:  |   : :|
Db     1495 KTALDQAAAD-KKTQIEQTPNASQQEINDAKQEVDTELNQAKTNIDQSSTDEYVDNAVKE 1553

Qy      789 ------SIKKFIPGRR--------------------------------------KKRADG 804
                  ::|  |      ::                                      |:  ||
Db     1554 GKAKINAVKTFSEYKKDALAKIEAAYNAKVTEADNSNASTSSEIAEAKQKLAELKQTADQ 1613

Qy      805 KQEQATVEDSGPVEINEDDPNV-----------PAVVPLSEYNAVEREKMEAQGNTELPQ 853
              |||  :|    |:|:  |  |:        |      |    |    ::  :  |  |      :
Db     1614 NVNQATSKDDIEVQIHNDLDNINDYTIPTGKKESATTDLYAYADQKKNNISADTNATQDE 1673

Qy      854 LLGAVYVSEELSKTLVHTVS---------------VAVIDGTRAVTSVEERSPSWISASV 898
              |:   ::   :|  : :::                |  ||  :    :|: ::    ||
Db     1674 KQQAIKQVDQNVQTALESINNGVDNGDVDDALTQGKAAIDAIQVDATVKPKANQVIDAKA 1733

Qy      899 TEPLE---------HTAGEAMPPVEEVTE--KDIIAEETPV--LTQTLPEGKDAHDDMV 944
              |   |                    ||:   ::::|:   |    |  : |     :  :    :| :| |::
Db     1734 EETKESIDQSDQLTAEEKTEALAMIKQITDQAKQGITDATTAEVEKAKAQGLEAFDNI- 1792

Qy      945 TSEVDFTSEAVTATETSEALRTEEVTEASGAEETTD-----------MVSAVSQLTDSP 992
              ::|  :     |    |    |            |:  ||:                :      :  ::|
Db     1793 --QIDSTEKQKAIEELETALDQIEAGVNVDADATTEEKEAFTNALEDILSKATEDISDQT 1850

Qy      993 DTTEEATPVQEVESGVLDTEEEERQTQAI----LQAVADKVKEESQV-------PATQTV 1041
               | |            |::    |:  :   : :|                 :  :  :     :  :  :
Db     1851 TNAEIAT----VKNSALEQLKAQRINPVVKKNALEAIREVVNKQIEIIKNADADASAKEI 1906

Qy     1042 QRTG-----SKALEKVEEVEEDSEVLASEKEKDVMPK----GPVQEAGAEHLAQGSE--- 1089
             ||          : :|:: : ::||   :|  :     :|              |   :   |   ||:
Db     1907 ARTDLGRYFDRFADKLDKTQTNTEV--AELQNVTIPAIEAIVPQNDPNANDTNSGSDNND 1964

Qy     1090 -----TGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTP 1144
                  ||||:    |  ||     |:                               ::|  |  |::: : :|
Db     1965 ATANSNANATPENTGQPNVTESTDNANA----------------DTSSTTTNNQNDAATG 2008

Qy     1145 LADSDTADGTQQDETID--SQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNN----VPA 1198
              : :|: :   |:  |       :: ::|      :    |     |:: |     ||   ||        ||
Db     2009 ETTATSANSSATDDANDKPQANNNSSADTSTNSPTMDNDVTSKPEVEST--NNGTTDKPA 2066

Qy     1199 QEEHGEEPGRDVLE------------PTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK 1246
              |    |                 ||:   ||    |    |    |    |    |     |
Db     2067 TEADNATPAESATNNNSTTTATNENAPTESTATA---PTTAST--GAESSADSKDNASVN 2121

Qy     1247 E-EQEVFVHSG------------PNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEG 1293
              :  :|     |::                  | |:   |                  |   :   :  |  :   |
Db     2122 DSKQNAEVNNSAESQSTNGKVAQPKSENKAKAEKDGRDSTNQSMVESTTETLPSADITEP 2181
```

```
Qy   1294 EMETDVEKEKRETKPEQVSEE---GEQETAAPEHEGTYG--------KPVLTLDMPSSER 1342
          :: ::  |:| |:    |    |    :    |: | : : |         ||  :   ||  :
Db   2182 KVSSNTSKDKEESTTSQTDAEQHNSDTNVASNEADKSEGNVDTDVSNKP--STSKPSEAK 2239

Qy   1343 GKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVG 1402
           ||         |  |  ||       :    ::  ||:   |  :              |
Db   2240 DKA-------TSTEDSQKADMATADTKDNQASIGATADVNNKATQN-----------DG 2280

Qy   1403 AHLLPAEKSSATGGHWTLQHAEDTVPLGPES-QAESIPIIVTPAPESTLHPDLQGEISAS 1461
           |:  ||                  ||  |   |    :  :  :   |    ::       ||:::
Db   2281 ANASPA-----------------TVSKGSNSANQDMLNVTKTKENKANAKSAQQGKVN-- 2321

Qy   1462 QRERSEEEDKPDAG 1475
           :  :  :  :   || |
Db   2322 -KPKQQAKTLPDTG 2334
```

RESULT 8
US-09-864-761-35612
; Sequence 35612, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 35612
; LENGTH: 1325
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AP000511.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 0.98
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EST_HUMAN HIT: AW867076.1, EVALUE 1.00e-103
; OTHER INFORMATION: EST_HUMAN HIT: AW867076.1, EVALUE 1.00e-103
; OTHER INFORMATION: SWISSPROT HIT: Q02817, EVALUE 5.00e-63
US-09-864-761-35612

```
  Query Match             4.8%;  Score 388;  DB 10;  Length 1325;
  Best Local Similarity   20.3%;  Pred. No. 6.6e-10;
  Matches  326;  Conservative  191;  Mismatches  662;  Indels  428;  Gaps   68;

Qy     53 QKNGQLSSVNGV--------------------------AEQGDVHVQEEN-QEGQEEEV 84
          || |:||: : :                            :  ::| |: :    :  :
Db      8 QKTGELSTSDHIFPLTPGLVYSIPFDHIVLHSGQRPPELPKSTEIHEQKRHCNTTRHSKP 67

Qy     85 VDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQP 144
           |:  |    |: :  |    :    |  |: |: :  |::           |   |  :  | |
Db     68 TDKPTG--NSKTIDHKSSTDNHEAPPTSEENSSNQGKD---------PMIRN--QRSVDP 114

Qy    145 AES---QANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQ 201
          |:|         | | :         |    :|     ||      :| | |:            |
Db    115 ADSTTTHKESAGKKHITPAPKSKINCRKSTTGKS----TVTRKSDK---------TGRPL 161

Qy    202 EPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQ 261
           | |: |    :  |   |     ::      |  : :| ||  |               |    |
Db    162 EKSMSTL--DKTSTSSHKTTTSFHNSGNSQTKQKSTSFP-----------EKITAASKT 207

Qy    262 EKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAE 321
            : |  :||         :| ||  :|   |                |: :  ||    :  ::
Db    208 TYKTTGTPEESEKTEDSRTTVASDKLLT-------------KTTKNIQETISANELTQSL 254

Qy    322 KVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATE 381
               |   :|   :    ||| |:     :|: |    |    :    :  |  :: |  ||
Db    255 AEPTEHGGRTANENNTPSPAEPTENRERTAN-ENTTLSPAEPTENRERTANENTAP---- 309

Qy    382 VFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQ 441
           |       ::|: |  :  :         | |                       |: |     |
Db    310 -FPAGPTENREMTANENTTLFPAEPTEHG--------------------ERTANENTTPS 348

Qy    442 EAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSP 501
           |||  |            |: |              | | |         |    |    ||
Db    349 PAEPTEH-----------GERT------ANENTTP-------SPAEPTEHGERTPFAND- 383

Qy    502 LKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPE 561
             |  |||        :    : |:|      |          ||: |      :     | : | |
Db    384 --KTTSSS-----AESTEHGERTPLANENTTPSPAEPTENRERTANENTTPSPAGPTENR 436

Qy    562 ETTCLEK---GPLEAPQDGEA---EEGTTSDGEKKREG-ITPWASFKKMVTPKKRVRRPS 614
          |||  ||     |:|  :: |     |: |  |   |     |  ||:|: |              |
Db    437 ETTANEKTTLSPVEPTENRETTANEKTTPSPAEPTENGQRTPFANEK---------TTSS 487

Qy    615 ESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEE--QKPEEPKRRVDTSVSWEALICV 672
           :: |  |:   |  ::|  |     ::  :|  |:    | ||     | :
Db    488 SAEPTEHGERTPLANENTTPSPAEPTENRERTANEKTTPSPAEPTENGDRT--------- 538

Qy    673 GSSKKRARKASSSDDEGGPRT---------LGGDSHRAEEASKDKEAGTDAVPASTQEQD 723
           : ::   : :    | | ||                  :     || :           |   ||   |
Db    539 PLANEKTTPSLAEPTENGQRTPFANEKTTSSSAEPTEHEERTPLANENTTPSPAEPTENR 598

Qy    724 Q--AQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEP 781
           :   |    :::| ||| |:|   ::   |   |   |  : ::::  |    | :  :      |:  ||
Db    599 ERTANENTTPSPAG-PTENREMTANE--KTTLFPAEPTENR-ERTANEKTT---SSPAEP 651

Qy    782 SREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVE-R 840
```

```
Db    652 TEN----------GQRTPFAN---EKTTSSPAEPTEHGERTPLANENTTLSPAEPTENR 697

Qy    841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Db    698 ERTANEKTTPFP---------------------AEPTENRERTANENTTP-----SPAQ 730

Qy    901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Db    731 PTEN--GDRTPLANEKT--------TPSLAEPTENGK---------RTPFANEKTT---S 768

Qy    961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Db    769 SSAEPTEHAERTPLANENT---------TSSP-----AEP----------TENRERTAN- 803

Qy   1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKV----EEVEEDSEVLASEKEKDVMPKGPV 1076

Db    804 ---------EKTTQFPAEPTENRE-STANEKTTPFPAEPTENREWTANE-NTTLSPAEPT 852

Qy   1077 QEAGAEHL---------AQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVA 1127

Db    853 EHEEMTPLANEKTTLSPAEPTENGERTP------------------------FTNEKTT 887

Qy   1128 PESSETLTDSETNGSTPLADSDT----ADGTQQDETIDSQDSKATAA--VRQSQVTEEEA 1181

Db    888 PSSAE---PTEHGERTPLANEITTPSRAEPTEHGERIANEKATPSPAKPTEHGETTVNED 944

Qy   1182 ATAQKEEPSTLPNNVP-AQE----------EHGEEPGRDVLEPTQQELTAAAVPVLAKTE 1230

Db    945 TTPSSAEPTENGERTPLANENTTTSPTESTEHGERTANEKTTPSPAEPTE---------- 994

Qy   1231 VGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSL 1290

Db    995 ----------HGERTPSANEKTIPSPAKPTEHEEMTPSA----------NENTTPSPVKP 1034

Qy   1291 EEGEMETDVEKEKRETKPEQVSEEGEQETAAPEH-EGTYGKPVLTLDMPSSERGKALGSL 1349

Db   1035 TEHGEKTTLANEKITLSPEGPTEHGAKTTSANEKITPSLAKPTEHGERTTSPNDKITSS- 1093

Qy   1350 GGSPSLPDQDKAGCIEVQVQSLDTTV---TQTAEAVEKVIETVVISETGESPECVGAHLL 1406

Db   1094 -AAESTEHRDRATSANVITPAPAEPIKHAKRTTLAHEKM--TQVTEKSTEHPEKTTS--- 1147

Qy   1407 PAEKSSATGGHWTLQHAEDTV-------PLGPESQAESI------------PIIVTPAPE 1447

Db   1148 TTEKTTRTPEKPTL-YSEKTICTKGKNTPV-PEKPTENLGNTTLTTETIKAPVKSTENPE 1205

Qy   1448 STLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIV 1507

Db   1206 KT--------------AAVTKTIKPSVKVTGDKSLTTTSSHLNKTE---VTHQVPTGSFT 1248

Qy   1508 LNVIQTAVDQFARTETAPETHAY----DSQTQVPACRLDSREPNRCW 1550

Db   1249 LITSRTKLSSITSEATGNESHPYLNKDGSQKGIHAGQMGENDSFPAW 1295
```

RESULT 9
US-09-971-536-70
; Sequence 70, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christensson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Methods
; TITLE OF INVENTION: Using Them
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; PRIOR FILING DATE: 2001-08-08
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 70
; LENGTH: 1879
; TYPE: PRT
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-70

```
  Query Match             4.7%;  Score 382;  DB 9;  Length 1879;
  Best Local Similarity   19.1%;  Pred. No. 1.8e-09;
  Matches  324;  Conservative  302;  Mismatches  726;  Indels  342;  Gaps   70;

Qy     20 TPSELVLSGHGPAAEASGAAGDPADADPATKLPQ-KNGQLSSVNGVAEQG---DVHVQEE 75

Db    124 TATEKAKQKDGVAAEATKAKAAIDQAQTETGVQQARDAGIAAIDAQHQPGTGLNVRREEA 183

Qy     76 NQEGQEEEV-----VDEDVGQRESEDVREKDRVEEMAANS-TAVEDITKDGQEETSEIIE 129

Db    184 KQAIDAEAAKVTAEIEQDSTLATSEKAAQKQGVADEAAKAKTAID------QAQTIEAID 237

Qy    130 Q----------------------------IPASENNVEEMV-QPAESQANDVGFKKVFK 159

Db    238 KAKDDGIKAIDAQHKQGADFDTRKAQAKDAIDAEAAKVKDAIDQDPTLTAKDKTAQK--Q 295

Qy    160 FVGFKFTVKKDKNEKSDTVQLLTVKKDEG-EGAEASVGAGDH---QEPSVETAVGESASK 215

Db    296 GVGDEATKAKTAIDQAKTIDGVIQAKDDGIKAIDAQHQAGTDLATRKDSAKQAIDAEAAK 355

Qy    216 ------ESELKQSTEKQ------------EGTLKQEQSSTEIPLQAESDQAAEEEAKDE 256

Db    356 ITDAINQDDTLTSTEKDAQKQAVADEAAKAKAAIDQAQNADAI-LQAQADGIKAIDAKHQ 414

Qy    257 GEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRK 316

Db    415 IGADLDTQKTKAKQA----IDKEAA----KVLT------------AIEQDPTLTSAEKK 453

Qy    317 EQEAEKVDEEEKEKTEPASEEQEPAEDTDQ-ARLSAD-------YEKVELPLEDQVGDLE 368

Db    454 AQKQGVADETAKAKTAIDS-----ARNADEIAKAQADGIKAIDAQHRLGMDLAKRKTDAQ 508

Qy    369 ASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQ--GGGGEAEGGVVVE---- 422

Db    509 AAIDAEAAKVGEAIDQDPTLTSQEKAAQKQTFAAEATKAKDTIAKAQDADGVIQAEKAGI 568

Qy    423 -------GTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQ---LTDLSPEE 472

Db    569 QAIDDGHQSGALLDTRKVDAKKAIDAEAAKINDAIDQDVTLTSAEKATQKQKVTDEAVKA 628

Qy    473 KTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPG 532

Db    629 KT-------AIDAAKNADTVDQAKASGIQAIDAVHQSG-TLLDTRKQDAKK--AIDAEAV 678

Qy    533 EY-----QHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDG 587

Db    679 KVIAAIGQDVTLTQAEKLTQQQAVADAATQ---------AKAAIDAAKNADAVDQAKADG 729

Qy    588 EK-----KREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSA-----TLSSTDSTV 637

Db    730 IKAIDAQHQAGL---ALNERKEAAKKLI---------AETADKVQAAIGQDVTLTATQKAV 778

Qy    638 SEMQDEVKTVGEEQ--------------KPEEPKRRVDTSVSWEALICVGSSKKRARKAS 683

Db    779 QRQAITVEVTKANQAIDAAGNADAVDQAKNAGVKAIYDQHQSGQAL----ADRKRDAK-Q 833
```

```
Qy     684 SSDDEGGPRTLGGDSH---RAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEG 740
             : | |       |             | | :  |:|   |      : :   | |: : : : |        :
Db     834 AIDAEAAKETAAIDQDATLTANEKASQKQAVADEATKAKEAIDAAKQADAVDQA----KN 889

Qy     741 EGVSTWES---FKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGR 797
             :|:    ::       :  |  ||  :       ::|  |:    :::   |:   ::|:   :  :|
Db     890 DGIRAIDAQHHAGQAVADRKAA----AKQAIDAEAAKVTGNID--QDETLTATEK----- 938

Qy     798 RKKRADGKQEQATVEDSGPVEINEDDPNVPAV--VPLSEYNAVEREKMEAQG----NTEL 851
                  |  ||  ||  |:      |::    |   ||            |:: :      |          :
Db     939 ----AAQKQAVATEADNAKQAIDKGQ-NADAVDKAKTGGIKAIDAQHQSGQAIKARQNDA 993

Qy     852 PQLLGAVYVSEELSKTLVHTVSVAVID---GTRAVTSVEERSPSWISAS-VTEPLEHTAG 907
              | :  |    : :::| :        ::      :      :|||   | :: :  |  |:  |  : ::
Db     994 KQAIDA--EAAKVTKAIDQDPTLTAAEKKAQKQAVTDAETKAKAAIDATLVADAIDQALA 1051

Qy     908 EAMPPVEEVTEKDI-IAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRT 966
             :  :   ::     :    |  : ::          ||:      ||        |:    :|       :  ||  |      :|  :
Db    1052 DGIKTIDAQYQTGIALDKQKAAAKQTI----DAEAAKVSEAID---QDVTLTADQKATQK 1104

Qy     967 EEVTEASGAEETTDMVSAVSQLTDSPDTTE------EATPVQEVESGVLDTEEEERQTQA 1020
             : |          |:| |    :|:  | :|:       :            ||      |         :    ||      ::::|  :
Db    1105 QAV-----ADEATKAKAAIDQASDADAVIQATIDGIEAIDAQHQSATALD--KQKQQAKQ 1157

Qy    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVL--------ASEKEKDVMP 1072
              : |  |  || :           |  |     :       :     : | ::    :::               :  |  |  :
Db    1158 AIDAEAAKVSKAIDQDVTLTATQKADQKQAVIAEADKAKKLIDAAGNADGIKQAESDGIK 1217

Qy    1073 KGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSE 1132
                  |     :: : ||            :  |          |   :  ||  :||  ||                |:|                :
Db    1218 AIDAQHQSSQALADRKRDAK-TAIDAEAAKETAAIDHDATLTA------NEKA---SQKQ 1267

Qy    1133 TLTDSETNGSTPLADSDTADGTQQDET-----IDSQ--------DSKATAAVRQSQVTEE 1179
              :||    |           :    :     ||     |  :|           ||:|                 |  ||   |          ||  :
Db    1268 AVTDEATKAKKAIDAAKQADAVDQAKTDGIKAIDAQHHSGQALDDRKADA----KQVIDA 1323

Qy    1180 EAA--TAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEV 1237
             |||     ||     :: : ||            | :: |   |                                   ||   :     |:
Db    1324 EAAKVTAAIDQDNTLTKAQKAAQKQGVATEAD---------------KAKQAIDAAGDA 1367

Qy    1238 DWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMET 1297
              | :|    |       |  :              ::|    |     |  ||                 ::|: :|        ::   :   :
Db    1368 DAVDQAKTAGIQAI------DAQHKAGKTIDSRHDDAKQAIDEEAAKV----IKAIDQDP 1417

Qy    1298 DVEKEKRETKPEQVSEEGEQETAAPEHEG---------TYG--------KPVLTLDMPSS 1340
              :      ::|  :  :  |:  |  ::        | :       |                   ||                  |        |:|
Db    1418 TLTAAQKEAQKQAVATEADKAKKAIDAAGDADAVDQAKTAGIKAIDEQHKSGQTVDARKE 1477

Qy    1341 ERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPEC 1400
             :   ||: :    |  :       |         |:: :       |        |:   :|  |:        ::    :    :    :
Db    1478 DAKKAIDAEAGKVTDAIDHDATLTAAQKEAQKQAVADEADKAKKAIDAAGNADAIDQAKS 1537

Qy    1401 VGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISA 1460
              |      :   :          ::    :|            :        |          |    |                |:
Db    1538 AGIKAIDEQHKSGQ----SIDTRKDDAKKAIDGEVAKITDAIDHDPTLT-----DAEKAT 1588

Qy    1461 SQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNK---IVLNVIQTAVDQ 1517
              ::           |  ||           ||  |  :: |:::  ||   :  :: :  : ||:  :     |   :   :  |:|:
Db    1589 QKQAVIAEADKAKKAIDAAG-DADAVDQAQKAGIKAIDQQHKSGQALAIRKDAAKKAIDE 1647

Qy    1518 FARTETAPETHAYD 1531
                  |  |  :  |  |
Db    1648 ----EAAKVSEAID 1657
```

RESULT 10
US-09-815-242-5816
; Sequence 5816, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5816

Query Match 4.7%; Score 377; DB 10; Length 2478;
Best Local Similarity 18.3%; Pred. No. 4.1e-09;
Matches 305; Conservative 274; Mismatches 625; Indels 462; Gaps 68;

```
Qy      47 PATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEE-EVVDEDVGQRESEDVREKD---- 101
                |    |     |           :  : :|     :|        :::  |      |:|  :|  :|  :: :|: |:: ::: |
Db     681 PTVKPQAKQDIIQAVTTRKQQ----IKKSNASLQDEKDVANDKIGKIETKAIKDIDAATT 736

Qy     102 RVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFV 161
                :    |   : |: ||  :          |::            |:       :|:||          ||
Db     737 NAQVEAIKTKAINDINQTTPATTAK-----AAALEEFDEVVQAQIDQA----------- 779

Qy     162 GFKFTVKKDKNEKSDTVQLLTVKKDEG-EGAEASVGAGDHQEPSVETAVGESASKESELK 220
                               | :: :: :       |    |   :   ||:   ||            :|         |     ||       :
Db     780 --PLNPDTTNEEVAEAIERINAAKVSGVKAIEATTAQD-----LERVKNEEISKIENIT 832

Qy     221 QSTE---------KQEGTLKQEQSST-------EIPLQAESDQAAEEEAKDEG------- 257
                ||:                 ||    |  ::  |::|                |:      ||:|  ||   :  |:  :|
Db     833 DSTQTKMDAYNEVKQAATARKAQNATVSNATNEEV---AEAD-AAVDAAQKQGLHDIQVV 888

Qy     258 EEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKE 317
                : |||     |||               :|:        |                            :           |   :  :  ::|   |        ::
Db     889 KSKQEVADTKS--KVLDKINAIQT---------------QAKVKPAADTEVENAYNTRK 930

Qy     318 QE-----AEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSE 372
                ||          |          :|::        ||       :::||         :    |    |        ::|
Db     931 QEIQNSNASTTEEKQAAYTELDTKKQEARTNLDAANTNSD------------------- 970

Qy     373 EKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEK 432
                               |    |    :    |    :|   |         :      |      |      |              :|          :|        :|       |:
Db     971 ------VTTAKDNSIAAINQVQAATTKKSDAKAEIAQKA---SERKTAIEAMNDSTTEEQ 1021

Qy     433 LAEPQEVPQEAEPA----EELMKSREMCVSGGDHTQLTD------LSPEEKTLPKHPEGI 482
                |      :|  |        |           :         : ::          |:  :  |:                ::|:          |         :    |
Db    1022 QAAKDKVDQAVVTANADIDNAAANNDV-----DNAKTTNEATIAAITPDANVKPAAKQAI 1076

Qy     483 VSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESP 542
                 :|:       :||                                        |       |         :|:         |   :  ||
Db    1077 ADKVQ---AQE----------------------TAIDGNNGSTTEEKAAAKQQVQTEK- 1109

Qy     543 ESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKK 602
```

```
Db   1110 TTADAAIDAAHTNAEVEAAKKAAIAK--IEAIQPAT----TTKD-----------NAKE 1151

Qy    603 MVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDT 662
Db   1152 AIATKANERKTAIAQTQDITAEEIAAANADVDNAVTQANSNIEAANSQNDVDQAKTTGEN 1211

Qy    663 SVSWEALICVGSSKKRARKASSS-------DDEGGPRTLGGDSHRAEEASKDKEAGTD-- 713
Db   1212 SI--DQVTPTVNKKATARNEITAILNNKLQEIQATP-----DATDEEKQAADAEANTENG 1264

Qy    714 ----AVPAST--QEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPR--KKSKSKLE- 764
Db   1265 KANQAISAATTNAQVDEAKANA-----------------EAAINAVTPKVVKKQAAKDEI 1307

Qy    765 EKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDP 824
Db   1308 DQLQATQTNVINNDQNATTEEKEAAI--------------QQLATA-------VTDAKN 1345

Qy    825 NVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVT 884
Db   1346 NITAA---TDDNGVDQAKDAGKNSIQSTQPATAVKSNAKNDVDQAVTTQNQAIDNTTGAT 1402

Qy    885 SVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKD-AHDDM 943
Db   1403 TEE--------------KNAAKDLVLKAKEKAYQDILNAQT---TNDVTQIKDQAVADI 1444

Qy    944 VTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAV-----SQLTDSPDTTEEA 998
Db   1445 QGITADTTIKDVAKDELATKANEQKALIAQTADATTEEKEQANQQVDAQLTQGNQNIENA 1504

Qy    999 TPVQEVESGVLDTEEEERQTQAI--LQAVADKVKEESQVPATQTVQRTGSKALEKVEEVE 1056
Db   1505 QSIDDVNTA------KDNAIQAIDPIQASTD-VKTNARAELLTEMQ-------NKITEIL 1550

Qy   1057 EDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVI 1116
Db   1551 NNNETTNEEKGNDI---GPVRAAYEEGL-----------NNINAATTTGDVTTAKDTAVQ 1596

Qy   1117 KLQQLMEQAV-APESSETLTDSETNGSTPLADSDTADGTQQD-----ETIDSQDSKATAA 1170
Db   1597 KVQQLHANPVKKPAGKKELDQAAADKKTQI--EQTPNASQQEINDAKQEVDTELNQAKTN 1654

Qy   1171 VRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTE 1230
Db   1655 VDQSS-TNEYVDNAVKEGKAKI-NAVKTFSEYKKD------------------ALAKIE 1693

Qy   1231 VGQEGEVDWLDGEKV---------------------------KEEQEVFVHSG----- 1256
Db   1694 DAYNAKVNEADNSNASTSSEIAEAKQKLAELKQTADQNVNQATSKDDIEVQIHNDLDNIN 1753

Qy   1257 ----PNSQKAA-------------------DVTYDSEVMGVAGCQEKESTEVQSL---- 1288
Db   1754 DYTIPTGKKESATTDLYAYADQKKNNISADTNATQDEKQQAIKQVDQNVQTALESINNGV 1813

Qy   1289 -------SLEEGEMETDVEKEKRETKPE-----QVSEEGEQETAAPEHEGTYGKPVLTLD 1336
Db   1814 DNGDVDDALTQGKAAIDAIQVDATVKPKANQAIEVKAEDTKESIDQSDQ----------- 1862

Qy   1337 MPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEK-------VIETV 1389
Db   1863 LTAEEKTEALAMI---KQITDQAKQGI---------TDATTTAE-VEKAKAQGLEAFDNI 1909

Qy   1390 VISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG-----------PESQAESI 1438
Db   1910 QIDST--------------EKQKAIE---ELETALDQIEAGVNVNADATTEEKEAFTNAL 1952

Qy   1439 PIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEIL- 1497
Db   1953 EDILSKATEDISDQTTNAEIATVKNSALEQLKAQRINPEVKKNALEAIREVVNKQIEIIK 2012

Qy   1498 --ELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQ---VPA 1538
Db   2013 NADADASAKEIARTDLGRYFDRFADKLDKTQTNAEVAELQNVTIPA 2058
```

RESULT 11
US-09-815-242-12967
; Sequence 12967, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 12967
;   LENGTH: 2478
;   TYPE: PRT
;   ORGANISM: Staphylococcus aureus
US-09-815-242-12967

```
  Query Match             4.7%;  Score 377;  DB 10;  Length 2478;
  Best Local Similarity   18.3%;  Pred. No. 4.1e-09;
  Matches  305;  Conservative  274;  Mismatches  625;  Indels  462;  Gaps   68;

Qy     47 PATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEE-EVVDEDVGQRESEDVREKD---- 101
Db    681 PTVKPQAKQDIIQAVTTRKQQ----IKKSNASLQDEKDVANDKIGKIETKAIKDIDAATT 736

Qy    102 RVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFV 161
Db    737 NAQVEAIKTKAINDINQTTPATTAK-----AAALEEFDEVVQAQIDQA------------ 779

Qy    162 GFKFTVKKDKNEKSDTVQLLTVKKDEG-EGAEASVGAGDHQEPSVETAVGESASKESELK 220
Db    780 --PLNPDTTNEEVAEAIERINAAKVSGVKAIEATTTAQD-----LERVKNEEISKIENIT 832

Qy    221 QSTE---------KQEGTLKQEQSST-------EIPLQAESDQAAEEEAKDEG------- 257
Db    833 DSTQTKMDAYNEVKQAATARKAQNATVSNATNEEV---AEAD-AAVDAAQKQGLHDIQVV 888

Qy    258 EEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKE 317
Db    889 KSKQEVADTKS--KVLDKINAIQT----------------QAKVKPAADTEVENAYNTRK 930

Qy    318 QE-----AEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSE 372
Db    931 QEIQNSNASTTEEKQAAYTELDTKKQEARTNLDAANTNSD-------------------- 970

Qy    373 EKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEK 432
```

```
Db     971 ------VTTAKDNSIAAINQVQAATTKKSDAKAEIAQKA---SERKTAIEAMNDSTTEEQ 1021

Qy     433 LAEPQEVPQEAEPA----EELMKSREMCVSGGDHTQLTD------LSPEEKTLPKHPEGI 482

Db    1022 QAAKDKVDQAVVTANADIDNAAANNDV-----DNAKTTNEATIAAITPDANVKPAAKQAI 1076

Qy     483 VSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESP 542

Db    1077 ADKVQ---AQE---------------------TAIDGNNGSTTEEKAAAKQQVQTEK- 1109

Qy     543 ESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKK 602

Db    1110 TTADAAIDAAHTNAEVEAAKKAAIAK--IEAIQPAT----TTKD-----------NAKE 1151

Qy     603 MVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDT 662

Db    1152 AIATKANERKTAIAQTQDITAEEIAAANADVDNAVTQANSNIEAANSQNDVDQAKTTGEN 1211

Qy     663 SVSWEALICVGSSKKRARKASSS-------DDEGGPRTLGGDSHRAEEASKDKEAGTD-- 713

Db    1212 SI--DQVTPTVNKKATARNEITAILNNKLQEIQATP-----DATDEEKQAADAEANTENG 1264

Qy     714 ----AVPAST--QEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPR--KKSKSKLE- 764

Db    1265 KANQAISAATTNAQVDEAKANA----------------EAAINAVTPKVVKKQAAKDEI 1307

Qy     765 EKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDP 824

Db    1308 DQLQATQTNVINNDQNATTEEKEAAI---------------QQLATA-------VTDAKN 1345

Qy     825 NVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVT 884

Db    1346 NITAA---TDDNGVDQAKDAGKNSIQSTQPATAVKSNAKNDVDQAVTTQNQAIDNTTGAT 1402

Qy     885 SVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKD-AHDDM 943

Db    1403 TEE--------------KNAAKDLVLKAKEKAYQDILNAQT---TNDVTQIKDQAVADI 1444

Qy     944 VTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAV-----SQLTDSPDTTEEA 998

Db    1445 QGITADTTIKDVAKDELATKANEQKALIAQTADATTEEKEQANQQVDAQLTQGNQNIENA 1504

Qy     999 TPVQEVESGVLDTEEEERQTQAI--LQAVADKVKEESQVPATQTVQRTGSKALEKVEEVE 1056

Db    1505 QSIDDVNTA------KDNAIQAIDPIQASTD-VKTNARAELLTEMQ-------NKITEIL 1550

Qy    1057 EDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVI 1116

Db    1551 NNNETTNEEKGNDI---GPVRAAYEEGL-----------NNINAATTTGDVTTAKDTAVQ 1596

Qy    1117 KLQQLMEQAV-APESSETLTDSETNGSTPLADSDTADGTQQD-----ETIDSQDSKATAA 1170

Db    1597 KVQQLHANPVKKPAGKKELDQAAADKKTQI--EQTPNASQQEINDAKQEVDTELNQAKTN 1654

Qy    1171 VRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTE 1230

Db    1655 VDQSS-TNEYVDNAVKEGKAKI-NAVKTFSEYKKD-------------------ALAKIE 1693

Qy    1231 VGQEGEVDWLDGEKV-----------------------------KEEQEVFVHSG----- 1256

Db    1694 DAYNAKVNEADNSNASTSSEIAEAKQKLAELKQTADQNVNQATSKDDIEVQIHNDLDNIN 1753

Qy    1257 ----PNSQKAA-------------------DVTYDSEVMGVAGCQEKESTEVQSL---- 1288

Db    1754 DYTIPTGKKESATTDLYAYADQKKNNISADTNATQDEKQQAIKQVDQNVQTALESINNGV 1813

Qy    1289 -------SLEEGEMETDVEKEKRETKPE-----QVSEEGEQETAAPEHEGTYGKPVLTLD 1336

Db    1814 DNGDVDDALTQGKAAIDAIQVDATVKPKANQAIEVKAEDTKESIDQSDQ----------- 1862

Qy    1337 MPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEK-------VIETV 1389

Db    1863 LTAEEKTEALAMI---KQITDQAKQGI---------TDATTTAE-VEKAKAQGLEAFDNI 1909

Qy    1390 VISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG-----------PESQAESI 1438

Db    1910 QIDST--------------EKQKAIE---ELETALDQIEAGVNVNADATTEEKEAFTNAL 1952

Qy    1439 PIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEIL- 1497

Db    1953 EDILSKATEDISDQTTNAEIATVKNSALEQLKAQRINPEVKKNALEAIREVVNKQIEIIK 2012

Qy    1498 --ELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQ---VPA 1538

Db    2013 NADADASAKEIARTDLGRYFDRFADKLDKTQTNAEVAELQNVTIPA 2058
```

RESULT 12
US-09-919-172-98
; Sequence 98, Application US/09919172
; Patent No. US20020119463A1
; GENERAL INFORMATION:
; APPLICANT: Faris, Mary
; APPLICANT: Turner, Christopher M.
; TITLE OF INVENTION: PROSTATE CANCER MARKERS
; FILE REFERENCE: PA-0036 US
; CURRENT APPLICATION NUMBER: US/09/919,172
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: 60/222,469
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 102
; SOFTWARE: PERL Program
; SEQ ID NO 98
; LENGTH: 3256
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. US20020119463A1 2700132CD1
US-09-919-172-98

```
  Query Match             4.7%;  Score 377;  DB 10;  Length 3256;
  Best Local Similarity   20.2%;  Pred. No. 5.6e-09;
  Matches  387;  Conservative  241;  Mismatches  762;  Indels  524;  Gaps  83;

Qy      14 QPAGSDTPSELVLSGH---------------GPAAEASGAAGDPADADPATKLPQKNGQL 58

Db    1412 QTSGETTHTDKVPGGEDKSINAFRETAKQKLDPAASVTGSKRHP-------KTKEKAQPL 1464

Qy      59 SSVNGVAEQGDVHVQEENQEGQEE----------EVVDEDVGQR--ESEDVREKDRVEEM 106

Db    1465 EDLAGWKELFQTPVCTDKPTTHEKTTKIACRSQPDPVDTPTSSKPQSKRSLRKVDVEEEF 1524

Qy     107 AANSTAVEDITK---------DGQE----------ETSEIIEQIPASENNVEEMVQPAES 147

Db    1525 FALRKRTPSAGKAMHTPKPAVSGEKNIYAFMGTPVQKLDLTENLTGSKRRLQTPKEKAQA 1584

Qy     148 QANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEAS----------VGA 197

Db    1585 LEDLAGFKELFQTRG--HTEESMTNDKTAKVACKSSQPDLDKNPASSKRRLKTSLGKVGV 1642

Qy     198 GDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEG 257

Db    1643 KEEL-----LAVGKLTQTSGETTHTHTEPTGDGKSMKAFMESP-----KQILDSAASLTG 1692

Qy     258 EEKQEKEPTKSPESPSSPVN----SETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAE 313

Db    1693 SKRQLRTPKGKSEVPEDLAGFIELFQTPSHTKESMTNEKT---TKVSYRASQPDLVDTPT 1749

Qy     314 KRKEQ--EAEKVDEEEK----EKTEPAS------------EEQE-------PAEDTDQ- 346

Db    1750 SSKPQPKRSLRKADTEEEFLAFRKQTPSAGKAMHTPKPAVGEEKDINTFLGTPVQKLDQP 1809

Qy     347 -------ARLSADYEKVELPLEDQVGDLE---------ASSEEK------CAPLATEVFD 384

Db    1810 GNLPGSNRRLQTRKEKAQ-ALEELTGFRELFQTPCTDNPTTDEKTTKKILCKSPQSDPAD 1868
```

```
Qy     385 -------------EKMEAHQEVVA--EVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLP 429
                        :|  :  :|:|  ::   |    :      ||  :       ||    |
Db    1869 TPTNTKQRPKRSLKKADVEEEFLAFRKLTPSAGKAMHTPKAAVGEEKDINTFVGT----P 1924

Qy     430 PEKL----------AEPQEVPQEAEPAEELMKSREMCVSGGDHTQ-------LTDLS--- 469
            |||          ||   ::|:  |:|   :|:  :  | ||:        :|::|
Db    1925 VEKLDLLGNLPGSKRRPQTPKEKAKALEDLAGFKELFQTPG-HTEESMTDDKITEVSCKS 1983

Qy     470 --PEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPL---KKLFSSSGLKKLSGKKQKGKRG 524
              |:    |              ||::|:|:    :    :::       |  : |||  :  |
Db    1984 PQPDPVKTP-----------TSSKQRLKISLGKVGVKEEVLPVGKLTQTSGKTTQTHRE 2031

Qy     525 GGGDEEPGEYQHIHTESPESA--DEQKGESSASSPEEP-EETTCLE-----KGPLEAPQD 576
             ||   |:      || :         |        |   | ||   ||      |    : |
Db    2032 TAGD---GKSIKAFKESAKQMLDPANYGTGMERWPRTPKEEAQSLEDLAGFKELFQTPD- 2087

Qy     577 GEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRP----SESDKEEELEKVKSATLSS 632
             || ||  |      |      :|        |  | ||   |||        |  |||  :|  |  ::
Db    2088 -HTEESTTDDKTTKIACKSP--PPESMDTPTSTRRRPKTPLGKRDIVEELSALKQLTQTT 2144

Qy     633 -TD-------------STVSEMQDEVKTV-GEEQKPEEPKRRVDTSVSWEAL------- 669
             ||              |    :   |  : |  |:::|  ||  :        |
Db    2145 HTDKVPGDEDKGINVFRETAKQKLDPAASVTGSKRQPRTPKGKAQPLEDLAGLKELFQTP 2204

Qy     670 ICVG------SSKKRARKASSSDDEGGPRTLGGDSHRA------EEASKDKEAGTDAV-- 715
             ||         :  |  | ::    |   |  |      | |:          ||  |       |  :|
Db    2205 ICTDKPTTHEKTTKIACRSPQPDPVGTPTIFKPQSKRSLRKADVEEESLALRKRTPSVGK 2264

Qy     716 ------PASTQEQDQAQGSSSP-EPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAE 768
                  ||   |:|      :|  :      |       |      |       ||::|:::  ||: |
Db    2265 AMDTPKPAGGDEKDMKAFMGTPVQKLDLPGNLPGSKRWPQ-----TPKEKAQA-LEDLAG 2318

Qy     769 DSSVEQLSTEIEPSREESWVSI---------------KKFIPGRRKKRADGKQEQATVED 813
             : |       :|: :|      |                    |  |    |  ::||  ::|      :
Db    2319 FKELFQTPGTDKPTTDEKTTKIACKSPQPDPVDTPASTKQRPKRNLRKADVEEEFLALRK 2378

Qy     814 SGPVEINEDDPNVPAVVPLSEYNA-VER--EKMEAQGN-----------TELPQLLGAVY 859
              |      |||            |   ||      :|::    ||                  |   :    |
Db    2379 RTPSAGKAMDTPKPAVSDEKNINTFVETPVQKLDLLGNLPGSKRQPQTPKEKAEALEDLV 2438

Qy     860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919
             :||  :|   ||         |       :|  |    :||            :|          :         ::
Db    2439 GFKELFQTPGHTEESMTDD---KITEVSCKSP--------QPESFKTSRSSKQRLKIPLV 2487

Qy     920 DIIAEETPV----LTQTLPEGKDAHDD---------MVTSEVDFTSEAVTATETSEALR 965
              :  :|  |:        ||:|   |      |    :                |         |  :  |   :    ||
Db    2488 KVDMKEEPLAVSKLTRTSGETTQTHTEPTGDSKSIKAFKESPKQILDPAASVTGSRRQLR 2547

Qy     966 T--------------EEVTEASGAEE---TTDMVSAVSQLTDSPDTTEEATPVQEVESGV 1008
             |                    :|:   |   |   |    |   |   :   :      :    |:  |:  ||    :
Db    2548 TRKEKARALEDLVDFKELFSAPGHTEESMTIDKNTKIPCKSPPPELTDTATSTKRCPKTR 2607

Qy    1009 LDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEK 1068
             |    |  :|                  |    ||      |   |       |  :  ::  :::       :                     |:
Db    2608 LRKEVKEE-----LSAVERLTQTSGQSTHTHKEPASGDEGIKVLKQRAKKKPNPVEEEPS 2662

Qy    1069 DVMPKGPVQEA-GAEHLA---QGSETGQATPES--------------LEVPEVTADVDH 1109
              |:  |  ::|     ||   ||      :   |||      ||  ||                 |||  :  ||
Db    2663 RRRPRAPKEKAQPLEDLAGFTELSETSGHTQESLTAGKATKIPCESPPLEVVDTTASTKR 2722

Qy    1110 VATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTA------------------ 1151
              :|  |:|            |         :    ::|:|  |      ||   :  |
Db    2723 HLRTRVQKVQ-------VKEEPSAVKFTQTSGETTDADKEPAGEDKGIKALKESAKQTPA 2775

Qy    1152 ----------------DGTQQDETI-------------DSQDSKATAAVRQSQVTEEEAA 1182
                              :   |  |  |  :                    |  |  |     :|     |:  |
Db    2776 PAASVTGSRRRPRAPRESAQAIEDLAGFKDPAAGHTEESMTDDKTTKIPCKSSPELEDTA 2835

Qy    1183 TAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDG 1242
             |:  |   |   |    |     |   :||                          ||  |   |  :|       ||   ||
Db    2836 TSSKRRPRTRAQKVEVKEE------------------LLAVGKLTQT----SGETTHTDK 2873

Qy    1243 EKVKEEQEVFVHSGPNSQKAADVTYDSE-VMG-----------------VAGCQEKESTE 1284
             |  |  |  :           |   :|        |:|   |:|                 :|   ||    |
Db    2874 EPVGEGKGTKAFKQPAKRK-----LDAEDVIGSRRQPRAPKEKAQPLEDLASFQELSQTP 2928

Qy    1285 VQSLSLEEGEME--TDVEKEKRET-KPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSE 1341
              :     |     :     |     |:   ::  ||  ::|        :        ||:  |         |    |   |  |  |
Db    2929 GHTEELANGAADSFTSAPKQTPDSGKPLKIS---RRVLRAPKVE-PVGDVVSTRD-PVKS 2983

Qy    1342 RGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE-----AVEKVIETVVISETGE 1396
             :  |:   ||     |  ||   :         |            :|||    |         ||:::|
Db    2984 QSKSNTSL---PPLPFKRGGG--------KDGSVTGTKRLRCMPAPEEIVEE-------- 3024

Qy    1397 SPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQG 1456
                         |||  |                    :       |           :|  :  |:           |           :
Db    3025 ---------LPASKK------------QRVAPRARGKSSEPVVIMKRSLRTSAKRIEPAE 3063

Qy    1457 EISASQRERSEEEDK-PDAGPDADG-KESTAIEKVLKAEPEILELESKSNKIVLNVIQTA 1514
             |::::   :  ::||   |    |:  |:   |       :     :     :||   :||  |:        :  :|   :|
Db    3064 ELNSNDMKTNKEEHKLQDSVPENKGISLRSRRQNKTEAEQQITEVFVLAERIEIN----- 3118

Qy    1515 VDQFARTE-----TAPETHAY--DSQTQVPACRLDSREPNRCWTKMKDAKMKHP 1561
                    |  |        |:||           |         :    |     |              |   ||             :     :     |
Db    3119 -----RNEKKPMKTSPEMDIQNPDDGARKPIPRDKVTENKRCLRSARQNESSQP 3167
```

RESULT 13
US-09-764-176-7
; Sequence 7, Application US/09764176
; Patent No. US20020127553A1
; GENERAL INFORMATION:
; APPLICANT: NOTEBORN, Mathieu Hubertus Maria
; APPLICANT: DANEN-VAN OORSCHOT, Astrid Adriana Anna Maria
; APPLICANT: ROHN, Jennifer Leigh
; APPLICANT: WEISS, Bertram
; TITLE OF INVENTION: APOPTIN-ASSOCIATING PROTEIN
; FILE REFERENCE: 4725US
; CURRENT APPLICATION NUMBER: US/09/764,176
; CURRENT FILING DATE: 2001-01-17
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
;   LENGTH: 1400
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Description of Sequence: Amino acid sequence deduced from the nuc
;   OTHER INFORMATION: leic acid seqeuence of AAP-
US-09-764-176-7

```
  Query Match             4.6%;  Score 368;  DB 10;  Length 1400;
  Best Local Similarity   19.6%;  Pred. No. 5.2e-09;
  Matches  283;  Conservative  226;  Mismatches  495;  Indels  442;  Gaps   65;

Qy      95 EDVREKDRVEEMAANSTAVEDI--TKDGQEETSEIIEQIPASENNVEEMVQPAESQANDV 152
             ::::  |:  :  |  |::   ::  |    |||        ::  :        ::||    ::   :  |
Db      92 DNLKFKNIINEEDADTMRLQPIGRDKDGLMYWYQLDQ-----DHNVRMYIEEQDDQDGS- 145

Qy     153 GFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVK------KDEGEGAEASVGAGDHQEPSVE 206
                      :|   |  :::||  ::|:  ||    :              |:    :      :|          :  ||:|
Db     146 ---------SWKCIV-RNRNELAETLALLKAQIDPVLLKNSSQDNSS-----RESPSLE 190

Qy     207 TAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAE--EEAKDEGEEKQEKE 264
                 |    |||   :    |::|      ::      |   ||  |:  |:  |   ||    :  |::  |||
Db     191 ---DEETKKEEETPKQEEQKE---SEKMKSEEQPMDLENRSTANVLEETTVKKEKEDEKE 244

Qy     265 PTK----------SPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLET--- 311
                |                  ||:      :    |  :  |||:                |      |:  :   |   :|
Db     245 LVKLPVIVKLEKPLPENEEKKIIKEESDSFKENVK------PIKVEVKECRADPKDTKSS 298

Qy     312 AEKRKEQEAEKVD-----EEEKEKTEPASEEQEPAEDTDQA---------RLSADYE--- 354
               ||     ||  |:::        :       |   ||   ::||   |    ::       ||                 :||  |::
Db     299 MEKPVAQEPERIEFGGNIKSSHEITEKSTEETEKLKNDQQAKIPLKKREIKLSDDFDSPV 358
```

```
Qy    355 -----KVELPLEDQVGDLEASSEEKCAPLAT---------------EVFDEKMEAH----- 390
Db    359 KGPLCKSVTPTKEFLKDEIKQEEETCKRISTITALGHEGKQLVNGEVSDERVAPNFKTEP 418

Qy    391 -----QEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVE-GTG----ESLPPEK-------- 432
Db    419 IETKFYETKEESYSPSKDRNIITEGNGTESLNSVITSMKTGELEKETAPLRKDADSSISV 478

Qy    433 ---------LAEPQEVPQE-----AEPAEEL-----MKSREMCVSGGDH----------- 462
Db    479 LEIHSQKAQIEEPDPPEMETSLDSSEMAKDLSSKTALSSTESCTMKGEEKSPKTKKDKRP 538

Qy    463 -----------------TQLTDLSPEEKTLPK-----------------------HPE 480
Db    539 PILECLEKLEKSKKTFLDKDAQRLSPIPEEVPKSTLESEKPGSPEAAETSPPSNIIDHCE 598

Qy    481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTE 540
Db    599 KLASEKEVVECQSTSTVGGQSVKKV----DLETL--------------KEDSEFTKVEMD 640

Qy    541 SPESADEQKGESSASSPEEPEETTCLEKGPLE---------APQDGEAEEGTTSDGEKKR 591
Db    641 NLDNA-------QTSGIEEPSET----KGSMQKSKFKYKLVPEEETTASENTEITSERQK 689

Qy    592 EGI---TPWASFKKMVTPKKRVRRP----SESDKEEELEKV-----KSATLSSTDSTVSE 639
Db    690 EGIKLTIRISSRKKKPDSPPKVLEPENKQEKTEKEEEKTNVGRTLRRSPRISRPTAKVAE 749

Qy    640 MQDE----VKTVGEEQKPEE-------------PKRRVDTS------------VSWEALIC 671
Db    750 IRDQKADKKRGEGEDEVEEESTALQKTDKKEILKKSEKDTNSKVSKVKPKGKVRWT---- 805

Qy    672 VGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSP 731
Db    806 --GSRTRGRWKYSSNDE----SEGSGSEKSSAASEEEE---------EKESEEAILADDD 850

Qy    732 EP---AGSPSEGEGVSTWESFKR-----------LVTP-----------RKKSKSKLEEKA 767
Db    851 EPCKKCGLPNHPELILLCDSCDSGYHTACLRPPLMIIPDGEWFCPPCQHKLLCEKLEEQL 910

Qy    768 EDSSVEQLSTEIEPSREESWV----SIKKFIP--------GRRKKRADGKQEQATVEDSG 815
Db    911 QDLDVALKKKERAERRKERLVYVGISIENIIPPQEPDFSEDQEEKKKDSKKSKANLLERR 970

Qy    816 PV--------EINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKT 867
Db    971 STRTRKCISYRFDEFDEAID--------EAIEDDIKEADG---------GGVGRGKDISTI 1014

Qy    868 LVH---TVSVAVIDGTRAVTSVEERSPSWISASVTEP-------------LEHTAGEAMPP 912
Db   1015 TGHRGKDIS-TILDEERK----ENKRPQRAAAARRKKRRRLNDLDSDSNLDEEESEDEFK 1069

Qy    913 VEEVTEKD-IIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTE 971
Db   1070 ISDGSQDEFVVSDENPDESEEDPPSND------DSDTDFCSRRLRRHPSRPMRQSRRLRR 1123

Qy    972 ASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDT-----EEEERQTQAILQAVA 1026
Db   1124 KTPKKKYSD--------DDEEEESEENS--RDSESDFSDDFSDDFVETRRRRSRRNQKRQ 1173

Qy   1027 DKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPV-QEAGAEHLA 1085
Db   1174 INYKEDSESDGSQKSLRRG-KEIRRVHK----RRLSSSSESEESYLSKNSEDDELAKESKR 1228

Qy   1086 QGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPL 1145
Db   1229 SVRKRGRSTDEYSEADEEEEEEEG-------KPSRKRLHRIETDEEESCDNAHGDANQPA 1281

Qy   1146 ADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQK-----------EEPSTLPNN 1195
Db   1282 RDS-------QPRVLPSEQESTKKPYRIESDEEEDFENVGKVGSPLDYSLVDLPST---- 1330

Qy   1196 VPAQEEHGEEPGRDVL----EPTQQ------ELTAAAVPVLAKTEVGQEGEVDWLDGEKV 1245
Db   1331 ------NGQSPGKAIENLIGKPTEKSQTPKDNSTASASLASNGTSGGQEA------GAPE 1378

Qy   1246 KEEQEV 1251
Db   1379 EEEDEL 1384
```

RESULT 14
US-09-963-875-1
; Sequence 1, Application US/09963875
; Patent No. US20020164307A1
; GENERAL INFORMATION:
; APPLICANT: Massachusetts General Hospital
; TITLE OF INVENTION: Stem Cells of the Islets of Langerhans and Their Use in Treating
; TITLE OF INVENTION: Mellitus
; FILE REFERENCE: 17633/1235
; CURRENT APPLICATION NUMBER: US/09/963,875
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US60/169082
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 60/215109
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 60/238880
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: US 09/731261
; PRIOR FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 58
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-963-875-1

Query Match 4.4%; Score 354; DB 9; Length 1618;
Best Local Similarity 20.1%; Pred. No. 2.5e-08;
Matches 342; Conservative 224; Mismatches 623; Indels 510; Gaps 75;

```
Qy    183 VKKDEGEGAEASVGAGDHQEPSVETAVGESASKE-----------------SELKQS--T 223
Db     26 VKALEEQNELLSAGLGGLRRQSADTSWRAHADDELAALRALVDQRWREKHAAEVARDNLA 85

Qy    224 EKQEGT--------LKQEQSSTEI-----PLQAE-------SDQAAEEEAKDE------- 256
Db     86 EELEGVAGRCEQLRLARERTTEEVARNRRAVEAEKCARAWLSSQGAELERELEALRVAHE 145

Qy    257 ----GEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAG---------WRKKTSFKK 303
Db    146 EERVGLNAQAACAPRLPAPPRPPAPAPEVEELARRLGEAWRGAVRGYQERVAHMETSLDQ 205

Qy    304 SKEDDLETAEKRKE--QEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPLE 361
Db    206 TRERLARAVQGAREVRLELQQLQAERGGLLERRAALEQRLEGRWQERLRAT-EKFQLAVE 264

Qy    362 DQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVST------VEKTEEEQGGGG-- 413
Db    265 ----ALEQEKQGLQSQIAQVLEGRQQLAHLKMSLSLEVATYRTLLEAENSRLQTPGGGSK 320

Qy    414 ----------EAEGGVVVEG--TGESLP---PEKLAEPQEVPQEAE-PAEELMKSREMCV 457
Db    321 TSLSFQDPKLELQFPRTPEGRRLGSLLPVLSPTSL--PSPLPATLETPVPAFLKNQEFLQ 378

Qy    458 SGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGK 517
Db    379 ARTPTLASTPIPP----TPQAPSPAV---------DAEIRAQDAPLSLLQTQGGRKQAPEP 426

Qy    518 KQKGKRGG------GGDEEP-GEYQHIHT-ESPESADEQKGESSASSPEEPEETTCLEKG 569
Db    427 LRAEARVAIPASVLPGPEEPGGQRQEASTGQSPED------HASLAPPLSPDHSS----- 475

Qy    570 PLEAPQDGEAEEGTT-----SDGEKKREGITP---------WASFKKMVTPKKRVRRPSE 615
```

```
Db    476 -LEA-KDGESGGSRVFSICRGEGEGQIWGLVEKETAIEGKVVSSLQQEIWEEEDLNRKEI 533

Qy    616 SDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGE------EQKPEEPKRRVDTSVSWEAL 669

Db    534 QDSQVPLEK---ETLKSLG---EEIQESLKTLENQSHETLERENQECPRSLEEDL--ETL 585

Qy    670 ICVGSSKKRARKASSSDDEGGPR--------------TLGGDSHRAEEASKDKEA----- 710

Db    586 KSLEKENKRAIKGCGGSETSRKRGCRQLKPTGKEDTQTLQSLQKENQELMKSLEGNLETF 645

Qy    711 ---GTD--AVPASTQEQDQ---AQGSSSPEPAGSPSEGEGVS------------------ 744

Db    646 LFPGTENQELVSSLQENLESLTALEKENQEPLRSPEVGDEEALRPLTKENQEPLRSLEDE 705

Qy    745 TWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTE-------IEPSREESWVSIKKFIPGR 797

Db    706 NKEAFRSLEKENQEPLKTLEEE-DQSIVRPLETENHKSLRSLEEQDQETLRTLEK--ETQ 762

Qy    798 RKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGA 857

Db    763 QRRRSLGEQDQMTLRP-------------PEKVDLEPLKSLDQE---------------- 793

Qy    858 VYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVT 917

Db    794 ----------------------------------------IARPLENENQEFLKSLKEES 813

Qy    918 EKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEE 977

Db    814 VEAVKSLETEILESLKSAGQE------------NLETLKSPETQAPLWTPEEINKSGGNE 861

Qy    978 TTDMVSAVSQLTDSPDTTEEATPVQEVESGVLD----------------------TEEE 1014

Db    862 SSRKGNSRTTGVCGSEPRDIQTPGRG-ESGIIEISGSMEPGEFEISRGVDKESQRNLEEE 920

Qy   1015 ERQTQAILQAVADKVKEESQ-VPATQTVQRTGSKALEKVEEVEEDSEV-LASEKEKDVMP 1072

Db    921 ENLGKGEYQESLRSLEEEGQELPQSADVQR-WEDTVEKDQELAQESPPGMAGVENKD--- 976

Qy   1073 KGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVD-------HVATCQVIKLQQLMEQA 1125

Db    977 -----EAELNLREQDGFTGK--EEVVEQGELNATEEVWFPGEGHPENPEPKEQRGLVEGA 1029

Qy   1126 VAPESSETLTDSETNGS-----------------TPLADSDTADGTQQ---DETIDSQDS 1165

Db   1030 SVKGGAEGLQDPEGQSQQVGTPGLQAPQGLPEAIEPLVEDDVAPGGDQASPEVMLGSEPA 1089

Qy   1166 KATAAV--------------RQSQVTEEEAATAQKEEPSTLPNNVPAQE----------- 1200

Db   1090 MGESAAGAEPGLGQGVGGLGDPGHLTREEVMEPPLEEESLEAKRVQGLEGPRKDLEEAGG 1149

Qy   1201 ---EHGEEPG--RDVLEPTQQ-----ELTA-----AAVPVLAKTEVGQEGEVDWLDGEKV 1245

Db   1150 LGTEFSELPGKSRDPWEPPREGREESEAEAPRGAEEAFPAETLGHTGSDAPSPWPLGSEE 1209

Qy   1246 KEEQ------------EVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEG 1293

Db   1210 AEEDVPPVLVSPSPTYTPILEDAPGLQPQAEGSQEAS-WGVQGRAEAGKVESEQEELGSG 1268

Qy   1294 EMETDVEKEKRETKPEQVSEEGEQETAAPEHE--GTYGK--------PVLTLDMPSSERG 1343

Db   1269 EIPEGLQEEGEESREE--SEEDELGETLPDSTPLGFYLRSPTSPRWTPLESRGHPLKETG 1326

Qy   1344 K------ALGSLG-GSPSLPDQDKAG---C-----IEVQVQSLDTTVT-------QTAEA 1381

Db   1327 KEGWDPAVLASEGLEEPSEKEEGEEGEEECGRDSDLSEEFEDLGTEAPFLPGVPGEVAEP 1386

Qy   1382 VEKVIETVV----------------ISETGESPE------------------CVGAHLLP 1407

Db   1387 LGQVPQLLLDPAAWDRDGESDGFADEEESGEEGEEDQEEGREPGAGRWGPGSSVGS--LQ 1444

Qy   1408 AEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQ--GEISASQRER 1465

Db   1445 ALSSSQRGEFLESDSVSVSVPWD-----DSLRGAVAGAPKTALETESQDSAEPSGSEEES 1499

Qy   1466 S----EEEDK-----------PDAGPDADGKESTAIEKVLKAEPEILELESKSNKI---V 1507

Db   1500 DPVSLEREDKVPGPLEIPSGMEDAGPGAD---------IIGVNGQGPNLEGKSQHVNGGV 1550

Qy   1508 LNVIQTAVDQFARTETAPE 1526

Db   1551 MNGLEQSEESGARNALVSE 1569
```

RESULT 15
US-09-864-761-36182
; Sequence 36182, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 36182
;   LENGTH: 617
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AC005529.7
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 0.94
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.1

```
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 5.6
;   OTHER INFORMATION: SWISSPROT HIT: P12036, EVALUE 2.00e-33
;   OTHER INFORMATION: EST_HUMAN HIT: BF445754.1, EVALUE 3.00e-22
US-09-864-761-36182

  Query Match            4.4%;  Score 351.5;  DB 10;  Length 617;
  Best Local Similarity  23.4%;  Pred. No. 1e-08;
  Matches  175;  Conservative  107;  Mismatches  258;  Indels  207;  Gaps   30;

Qy      28 GHGPAAEASGAAGDPADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDE 87
            | ||    |    | |      :| :   |: :    |    :|:  | |:|: :| |     | |
Db      11 GFGP-IPFSLPEGLPKIPSVSTHIKVKSEEKIKVVEKSEKETVIVEEQTEETQ----VTE 65

Qy      88 DVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAES 147
            :| :  | :: :|::  ||          |:   : |:|||                 |||
Db      66 EVTEEEEKEAKEEEGKEEEGGE----EEEAEGGEEETKS----------------PPAEE 105

Qy     148 QANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVET 207
            |:                    :| ||        ||::     |||         : |  |
Db     106 AAS------------------PEKEAKS------PVKEEAKSPAEA-------KSPEKEE 134

Qy     208 AVGESASKESELKQSTEKQEGTLKQEQSSTE-----IPLQAESDQAAEEEAKDEGEEKQE 262
            |   :   |   |   :|   |:|         |  | |      | : :|: : |: ||:| :   |
Db     135 AKSPAEVKSPEKAKSPAKEEAKSPPEAKSPEKEEAKSPAEVKSPEKAKSPAKEEAKSPAE 194

Qy     263 -----------KEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLET 311
                       ||  |||    |||  |   |               |  :      |    ::: ::
Db     195 AKSPEKAKSPVKEEAKSPAEAKSPVKEEAKSP---------AEVKSPEKAKSPTKEEAKS 245

Qy     312 AEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSA-DYEKVELPLEDQVGDLEAS 370
            ||  |   | |:     || |:   :| : |  :        :   |    || : |:::     ||
Db     246 PEKAKSPEKEEAKSPEKAKSPVKAEAKSPEKAKSPVKAEAKSPEKAKSPVKE-----EAK 300

Qy     371 SEEKC-APLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLP 429
            | ||  :|:         | ||      |:  |      || |          :|:  |        |:
Db     301 SPEKAKSPV-------KEEAKSPEKAKSPVKEEAKTPE------KAKSPV----KEEAKS 343

Qy     430 PEKLAEPQEV------------------------PQEAE-PAEELMKSREMCVSGGDHT 463
            |||    |::                           |::|:  |  :|  :|| |    |
Db     344 PEKAKSPEKAKTLDVKSPEAKTPAKEEARSPADKFPEKAKSPVKEEVKSPEKAKS----P 399

Qy     464 QLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKR 523
                 |      || :||   | : | |:         : :||:  | |        :      |      |  : |:
Db     400 LKADAKAPEKEIPKKEE-VKSPVKEEEKPQEVKVKEPPKK----AEEEKAPATPKTEEKK 454

Qy     524 GGGGDEEPGEYQHIHTESPESADEQKGESSASSPEE------PEETTCLEKGP------- 570
               :| |  :         |:|:     |:| |   :     |:|           ||        :| |
Db     455 DSKKEEAPKK------EAPKPKVEEKKEPAVEKPKESKVEAKKEEAEDKKKVPTPEKEAP 508

Qy     571 --LEAPQDGEAEEGT-----TSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELE 623
              :|   :| : :|  |              |   | :|     |    : ||     |:|:   :    :: |  ||
Db     509 AKVEVKEDAKPKEKTEVAKKEPDDAKAKEPSKP--AEKKEAAPEKKDTKEEKAKKPEEKP 566

Qy     624 KVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKAS 683
            |           | :: ::: ||: :|   | :||                       ::| |:|
Db     567 K-----------TEAKAKEDDKTLSKE--PSKPK-----------------AEKAEKSS 595

Qy     684 SSDDEGGPRTLGGDSHRAEEASKDKEA 710
            |:| :             ||    |:|::|| |
Db     596 STDQK--------DSKPPEKATEDKAA 614

Search completed: December 13, 2002, 00:28:36
Job time : 248 secs
```

OM protein - protein search, using sw model

```
Run on:         December 13, 2002, 00:06:25 ; Search time 59 Seconds
                                              (without alignments)
                                              2600.518 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters:      283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       PIR_73:*
                 1:  pir1:*
                 2:  pir2:*
                 3:  pir3:*
                 4:  pir4:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	5421.5	67.2	1346	2	A57376	probable regulator
2	4083.5	50.6	1684	2	JW0057	gravin - human
3	585.5	7.3	1829	2	T24583	hypothetical prote
4	574	7.1	5327	2	T13564	microtubule-associ
5	501	6.2	2464	1	QRMSP1	microtubule-associ
6	497.5	6.2	3488	2	T34418	hypothetical prote
7	493	6.1	3924	2	S37431	ankyrin 2, neurona
8	489	6.1	6642	2	T29757	protein UNC-89 - C
9	488	6.0	5170	2	T15348	hypothetical prote
10	485	6.0	2484	2	T26216	hypothetical prote
11	485	6.0	2607	2	T26215	hypothetical prote
12	483.5	6.0	1948	2	S00485	gene 11-1 protein
13	470	5.8	2364	2	A56577	microtubule-associ
14	466.5	5.8	1871	2	D96796	probable heat shoc
15	456.5	5.7	1804	2	T34518	nestin - golden ha
16	438	5.4	1128	2	T30296	R27-2 protein - Tr
17	434	5.4	1526	2	A45605	mature-parasite-in
18	433.5	5.4	2774	2	A43359	microtubule-associ
19	425	5.3	1110	2	I51116	NF-180 - sea lampr
20	419	5.2	3187	2	JC5837	364K Golgi complex
21	411.5	5.1	3562	2	A47171	chondroitin sulfat
22	408.5	5.1	4687	1	A39638	plectin - rat
23	407.5	5.0	2139	2	T18296	myosin heavy chain
24	403.5	5.0	1909	2	A45592	liver stage antige
25	402.5	5.0	7962	2	I38346	elastic titin - hu
26	402	5.0	2218	2	B84683	hypothetical prote
27	401.5	5.0	4574	2	G02520	plectin - human
28	400.5	5.0	4684	2	A59404	plectin [imported]
29	400	5.0	1830	2	A37981	microtubule-associ
30	400	5.0	3225	2	I52300	giantin - human
31	397	4.9	3507	2	T34513	hypothetical prote
32	395.5	4.9	1825	2	S13507	microtubule-associ
33	393	4.9	3259	1	A56539	giantin - human
34	392.5	4.9	1822	2	S33441	EF protein - Strep
35	389	4.8	1262	2	T22523	hypothetical prote
36	389	4.8	1621	2	A82255	hypothetical prote
37	388	4.8	1805	2	A34736	nestin - rat
38	385.5	4.8	1087	1	QFMSH	neurofilament trip
39	382.5	4.7	1320	2	JC5630	TCOF1 protein - mo
40	382	4.7	2116	2	A26655	myosin heavy chain
41	380	4.7	1558	2	B71603	RESA-H3 antigen PF
42	379	4.7	2094	2	S33124	tpr protein - huma
43	379	4.7	2722	2	T20532	hypothetical prote
44	379	4.7	2738	2	E88320	protein F07A11.6 [
45	377	4.7	1634	2	T26517	hypothetical prote

ALIGNMENTS

```
RESULT 1
A57376
probable regulatory protein 322 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 08-Feb-1996 #sequence_revision 08-Feb-1996 #text_change 08-Feb-1996
C;Accession: A57376
R;Lin, X.; Nelson, P.J.; Frankfort, B.; Tombler, E.; Johnson, R.; Gelman, I.H.
Mol. Cell. Biol. 15, 2754-2762, 1995
A;Title: Isolation and characterization of a novel mitogenic regulatory gene, 322, whic
A;Reference number: A57376; MUID:95257957; PMID:7739556
A;Accession: A57376
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1346 <LIN>
A;Cross-references: GB:U23146
C;Genetics:
A;Gene: 322

  Query Match             67.2%;  Score 5421.5;  DB 2;  Length 1346;
  Best Local Similarity   91.6%;  Pred. No. 5.5e-187;
  Matches 1098;  Conservative   24;  Mismatches   62;  Indels   15;  Gaps    7;

Qy        387 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPA 60

Qy        447 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 EELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLF 120

Qy        507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 180

Qy        567 EKGPLEAPQDGEAEEGTTSDGEKKRE----GITPWASFKKMVTPKKRVRRPSESDKEEEL 622
              |||||||: |:  :       ||||:      ||      :|     :: ||||||||||
Db        181 EKGPLEAPRMGKLRKELLR-GEKKRKDHSLGI-----LQKDGDTQETVRRPSESDKEEEL 234

Qy        623 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 682
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        235 EKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA 294

Qy        683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 742
              ||||  ||||||    ::    : : : |||||||||||||||||||||||||||||||
Db        295 SSSDIR-GPRTLGGGQSQSRGGQQRQRSRTDAVPASTQEQDQAQGSSSPEPAGSPSEGEG 353

Qy        743 VSTWESFKRLVTPRKKSKSKLEEK-AEDSSVEQLSTEIEPSREESWVS--IKKFIPGRRK 799
              |||||||||||||||||||||||| | : |     : :| |  :    ::   |    :
Db        354 VSTWESFKRLVTPRKKSKSKLEEKEAGRTLVVGAGCPLRSNRVEKNLGFPLRNSSPDGGR 413

Qy        800 KRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVY 859
              |   |:||||||||||||||:|:||||||||||:|||||||||||| |||   | |
```

```
Db     414 KGQMGRQEQATVEDSGPVEINEDEPDVPAVVPLSEYDAVEREKMEAQGNAELPSCWGCV- 472

Qy     860 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 919
Db     473 VSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEK 532

Qy     920 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 979
Db     533 DIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETT 592

Qy     980 DMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039
Db     593 DMVSAVSQLTDSPDTTEEATPVQEVEGGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 652

Qy    1040 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 1099
Db     653 TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLE 712

Qy    1100 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 1159
Db     713 VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDET 772

Qy    1160 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELT 1219
Db     773 IDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELA 832

Qy    1220 AAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
Db     833 AAAVPVWQKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 892

Qy    1280 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPS 1339
Db     893 KESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHERNYGKPVLTLDMPS 952

Qy    1340 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399
Db     953 SERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1012

Qy    1400 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1459
Db    1013 CVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEIS 1072

Qy    1460 ASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1519
Db    1073 ASQRERSEEEDKPDAGPDADGKESTAIDKVLKAEPEILELESKSNKIVLNVIQTAVDQFA 1132

Qy    1520 RTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAW 1578
Db    1133 RTETAPETHAYDSQTQVPAMRLDSREPNRCWTKMKVAKMKHPVPQPREDLQVLTVLEAW 1191
```

```
RESULT 2
JW0057
gravin - human
C;Species: Homo sapiens (man)
C;Date: 13-Jun-1998 #sequence_revision 10-Jul-1998 #text_change 08-Oct-1999
C;Accession: JW0057
R;Sato, N.; Kokame, K.; Shimokado, K.; Kato, H.; Miyata, T.
J. Biochem. 123, 1119-1126, 1998
A;Title: Changes of gene expression by lysophosphatidylcholine in vascular endothelial c
A;Reference number: JW0057; MUID:98269042; PMID:9604001
A;Accession: JW0057
A;Molecule type: mRNA
A;Residues: 1-1684 &lt;SAT&gt;
A;Cross-references: DDBJ:AB003476; NID:g2081606; PIDN:BAA19927.1; PID:d1020716; PID:g208
C;Comment: This protein regulates cell growth.
F;433-439/Region: nuclear location signal
F;522-527/Region: nuclear location signal
F;591-596/Region: nuclear location signal
F;671-676/Region: nuclear location signal

  Query Match             50.6%;  Score 4083.5;  DB 2;  Length 1684;
  Best Local Similarity   58.3%;  Pred. No. 5.6e-139;
  Matches  903;  Conservative  179;  Mismatches  378;  Indels   89;  Gaps   30;

Qy      89 VGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQ 148
Db       9 VGQRDSEDVSKRDSDKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELTQPTESQ 68

Qy     149 ANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETA 208
Db      69 ANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHKDPSL--G 123

Qy     209 VGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKS 268
Db     124 AGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKS 182

Qy     269 PESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEK 328
Db     183 AESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEED 242

Qy     329 EKTEPASEE---------QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLA 379
Db     243 GKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLA 302

Qy     380 TEVFDEKMEAHQ-EVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQE 438
Db     303 TEVFDEKIEVHQEEVVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVEMDA 354

Qy     439 VPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQ 498
Db     355 EPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQ 414

Qy     499 GSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPE 558
Db     415 GSPLKKLFTSTGLKKLSGKKQKGKR-GGDEESGEHTQVPADSPDSQEEQKGESSASSPE 473

Qy     559 EPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDK 618
Db     474 EPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDK 533

Qy     619 EEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKR 678
Db     534 EDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKR 593

Qy     679 ARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPS 738
Db     594 ARRGSSSDEEGGPKAMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPT 653

Qy     739 EGEGVSTWESFKRLVTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFI 794
Db     654 EGEGVSTWESFKRLVTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFI 713

Qy     795 PGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTEL 851
Db     714 PGRRKKRPDGKQEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKSAEQ 773

Qy     852 PQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMP 911
Db     774 PEQKAATEVSKELSESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAAL 833

Qy     912 PVEEVTEKDIIA-EETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVT 970
Db     834 LTEEVLEREVIAEEEPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGAEEGT 893

Qy     971 EASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK 1030
Db     894 EASAAEETTEMVSAVSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVK 953

Qy    1031 EESQVPAT-------QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEH 1083
Db     954 EESQLPGTGGPEDVLQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEP 1007

Qy    1084 LAQGSETGQATPESLE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSET 1133
Db    1008 FTQGKVVGQTTPESFEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVET 1067
```

```
Qy   1134 LTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLP 1193
Db   1068 PTDSETDGSTPVADFDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-P 1126

Qy   1194 NNVPAQEEHGEEPG-RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----E 1247
Db   1127 SSFVFQEETKEQSKMEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFE 1184

Qy   1248 EQEVFVHSG--PNSQKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVE 1300
Db   1185 GLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVE 1244

Qy   1301 KEKRETKPEQVSEEG-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQ 1358
Db   1245 REKTEAEPTHVNEEKLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQ 1304

Qy   1359 DKAGCIEVQVQSLDTTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGG 1416
Db   1305 EEAVCTKIQVQSSEASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNE 1364

Qy   1417 HWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGP 1476
Db   1365 DFAAHPGEDAVPTGPDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQ 1424

Qy   1477 DADGKESTAIEKVLKAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQ 1533
Db   1425 EV--KVSVAIEDL---EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQ 1479

Qy   1534 TQVPACRLDSREPNRCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
Db   1480 TQAHVIKADSQDAGQETEKEGEEPLASAQDETPITSAKEESESTAVGQA 1528
```

RESULT 3
T24583
hypothetical protein T06D8.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T24583
R;Palmer, S.
submitted to the EMBL Data Library, April 1995
A;Reference number: Z19909
A;Accession: T24583
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1829 <WIL>
A;Cross-references: EMBL:Z49130; PIDN:CAA88964.1; GSPDB:GN00020; CESP:T06D8.1
A;Experimental source: clone T06D8
C;Genetics:
A;Gene: CESP:T06D8.1
A;Map position: 2
A;Introns: 1391/3; 1432/3; 1470/3; 1505/1; 1520/1; 1616/1; 1644/1; 1687/3; 1742/1

```
  Query Match             7.3%;  Score 585.5;  DB 2;  Length 1829;
  Best Local Similarity   23.2%;  Pred. No. 1e-13;
  Matches  375;  Conservative  174;  Mismatches  662;  Indels  405;  Gaps   65;

Qy      2 GAGSSTE-----QRSPEQPAGSDTPSELVLSGHGPAAEASG-----AAGDPADADPAT-- 49
Db    106 GSGEETTVVAVVESSGEEPASSSTSVPTELSKDDQVTEASGEETTTAAATEASSEETTTS 165

Qy     50 KLPQKNGQLSSVNGVAE--------------QGDVHVQEENQEGQEEE----VVDEDVGQ 91
Db    166 AVTEGSGGEETTTSAVTEASSEATTTPAGTEASGEETTTSAVTEGSGGEETTVVAVVESSGE 225

Qy     92 RESE-------DVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMV-- 142
Db    226 EPASSSTSIPTELSKNDQVTEASGEETITAAATEASEETTSAVTEGSGEDTTVVAVVEL 285

Qy    143 ---QPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAG- 198
Db    286 SGEQPASSS-----------------TSIPTELSKDDQV-------TEASGEETTTAAAT 321

Qy    199 DHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGE 258
Db    322 EASEETTTSAVTEGSGEETTVVAVVESSGE--EPASSSTSIPTELSKDDQVTEAS---GE 376

Qy    259 EKQEKEPTK-SPESPSSPV---NSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEK 314
Db    377 ETTTAAATEASEETTTSAVTEGSGEDTTVVAVVESSGEQPASSSTSIPTELSKDDQVTEA 436

Qy    315 RKEQEAEKVDEEEKEKTEPASEEQEPAEDT----------DQARLSADYEKVELPLEDQV 364
Db    437 SGEETTTAAATEASEETTTSAVTEGSGEDTTVVAVVESSGEQPASSSTSIPTELSKDDQV 496

Qy    365 GDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGT 424
Db    497 --TEASGEETTTAAATEASEE-------------TTTSAVTE---GSGEETTVVAVVESS 538

Qy    425 GESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHT---QLTDLSPEEKTLPKHPEG 481
Db    539 GE----EPASSSTSIP------TELSKDDKVTEASGEETTTAAATDASSEETTTSAVTEG 588

Qy    482 IVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQ-HIHTE 540
Db    589 SGEETTVVAVVE--------------------------------SSDEEPASSSTSIPTE 616

Qy    541 SPESADEQKGESS------ASSPEEPEETTCLEKGPLEAPQDGEAEEGTT-----SDGEK 589
Db    617 L--SKDDQVTEASGEETTTAAATEASEETT------TSAVTEGSGEETTVVAVVESSGEE 668

Qy    590 KREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGE 649
Db    669 ------PASSSTSIPT------ELSKDDKVTEASGEETTTAAATDASSEETTTSAVTEGS 716

Qy    650 ----------EQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSH 699
Db    717 GEETTVVAVVESSDEEPASS-STSIPTEL----------------SKDDQVTEASGEETT 759

Qy    700 RAEEASKDKEAGTDAVPASTQEQDQAQG---SSSPEPAGSPSEGEGVSTWESFKRLVTPR 756
Db    760 TAAATEASEETTTSAVTEGSGEETTVVAVVESSGEEPASSSTS--------------IPT 805

Qy    757 KKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGP 816
Db    806 ELSKD--DQVTEASGEETTTAAATEASEETTTS-------------------AVTEGSG- 843

Qy    817 VEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAV 876
Db    844 -----EDTTVVAVVESS-------GEQPASSSTSIP---------TELSK---------- 872

Qy    877 IDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEG 936
Db    873 -DDQVTEASGEETT----TAAATEASEETTSA------VTEGS--GEETTVVAVVESSG 919

Qy    937 KD-----------AHDDMVTSEVDFTSEAVTATE-TSEALRTEEVTEASGAEETTDMVS 983
Db    920 EEPASSSTSIPTELSKDDQVTEASGEETTTAAATEASSEETTTSAVTEGSGEETTTSAVT 979

Qy    984 AVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK----EESQVPATQ 1039
Db    980 EGS----GEETTTSAVP-----------EGENSTTEAPAFVTGSEIEIPSSEESSSTTTH 1024

Qy   1040 -------TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEH------LAQ 1086
Db   1025 DPSIPVITPKPSVSSTIENVMS-KTSSEEAAEKKIIGEHQTGKDDDAGKEDEDNMPAFVT 1083

Qy   1087 GSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLA 1146
Db   1084 ANPAGTSTTESAENVTSTGEED-----ENIKMAKELGKQFAAD---------------LA 1123

Qy   1147 DSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEP 1206
Db   1124 KLAAKDGVNLTETADAKDSGETAHVEDEQVSSTESSIGSEETTTTV--NKETTEEHHEAS 1181
```

```
Qy   1207 GRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVT 1266
Db   1182 G--------EEDDAPAFVTGAPTDSTTEASVSTTSA--ITDETTSVAADESTSTSAGEVQ 1231

Qy   1267 YDSEVMGVA--GCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEH 1324
Db   1232 SSSAIIDSATVASEEQTSSEATSVIESSGE-EVTTTDENLVTSTVAQLEEGSGITAAESK 1290

Qy   1325 EGTYGKPVLTLDMPSSERGKALGSLGGSPS--LPDQDKAGCIEVQVQSLD----TTVTQT 1378
Db   1291 D----EDSVTTEATSQSTTVSESSDGSGESTVAPNDSETSTTESSQSTTDEGSGVTAAES 1346

Qy   1379 AEAVEKVIE--TVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTV---PLGPES 1433
Db   1347 KDEESSTTEAPAFVTSKTSGSEE---------DEEDSPDTHEFLTGIDETMFNKSLVPDT 1397

Qy   1434 QAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKV 1489
Db   1398 HREDLPNNVGFVPSSEPKPKNPDE---------EEEEEEDDGTKSDDYEDNVSKKI 1444
```

RESULT 4
T13564
microtubule-associated protein homolog - fruit fly (Drosophila melanogaster)
N;Alternate names: hypothetical protein EG:49E4.1
C;Species: Drosophila melanogaster
C;Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 17-Nov-2000
C;Accession: T13564
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of Drosophila melanogaster.
A;Reference number: Z17689
A;Accession: T13564
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5327 <SPA>
A;Cross-references: EMBL:AL031128; PIDN:CAA20006.1
C;Genetics:
A;Cross-references: FlyBase:FBgn0025392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1
C;Superfamily: Drosophila 576K microtubule-associated protein homolog

```
  Query Match             7.1%;  Score 574;  DB 2;  Length 5327;
  Best Local Similarity   21.6%;  Pred. No. 9.2e-13;
  Matches  415;  Conservative  285;  Mismatches  734;  Indels  484;  Gaps   82;

Qy      6 STEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPAD-------ADPATKLPQKNGQL 58
Db   2744 SVVESSKDDAEKSESRPESVIASGEPVPRESKSPLDSKDTSRPGSVVESVTAEDEKSEQQ 2803

Qy     59 SSVNGVAE--QGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMA----ANSTA 112
Db   2804 SRRESVAESVKADTKKDGKSQEASRPSSVDELL---KDDDEKQESRRQSITGSHKAMSTM 2860

Qy    113 VEDITKDGQEETSEIIEQIPASENNVEEMVQ----------------PAESQANDV--- 152
Db   2861 GDESPMDKADKSKE-----PSRPESVAESIKHENTKDEESPLGSRRDSVAESIKSDITKG 2915

Qy    153 -----GFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEP---- 203
Db   2916 EKSPLPSKEVSRPESVVGSIKDEKAESRRESVAESVKPESSKDATSAPPSKEHSRPESVL 2975

Qy    204 -----------SVETAVGESASKESELKQSTE-----------KQEGTLKQEQSSTEIPL 241
Db   2976 GSLKDEGDKTTSRRVSVADSIKDEKSLLVSQEASRPESEAESLKDAAAPSQETSRPESVT 3035

Qy    242 QAESD-------------QAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFF 288
Db   3036 ESVKDGKSPVASKEASRPASVAENAKDSADESKEQRPESLPQSKAGSIKDEKSP------ 3089

Qy    289 THGWAGWRKKTSFKKSKEDD--------------------------LETAEKRKEQ---- 318
Db   3090 ------LASKDEAEKSKEESRRESVAEQFPLVSKEVSRPASVAESVKDEAEKSKEESPLM 3143

Qy    319 ----------------EAEKVDEEEK-----EKTE-PASEEQEPAEDTDQARLSADYEKV 356
Db   3144 SKEASRPASVAGSVKDEAEKSKEESRRESVAEKSPLPSKEASRPASVAESVKDEADKSKE 3203

Qy    357 ELPLEDQVGDLEASSEEKCAP--LATEVFDE----KMEAHQEVVAEVH----------- 398
Db   3204 ESRRESGAEKSPLASKEASRPASVAESIKDEAEKSKEESRRESVAEKSPLPSKEASRPTS 3263

Qy    399 -----VSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKSR 453
Db   3264 VAKSVKDEAEKSKEESSRDSVAEKSPL--ASKEASRPASVAE--SVQDEAEKSKE--ESR 3317

Qy    454 EMCVSGGDHTQLTDLSPEEKTLPKH-PEGIVSEVEMLSSQERIK--VQGSPLKKLFSS-- 508
Db   3318 RESVA-----EKSPLAYKEASRPASVAESIKDEAEKSKEESRRESVAEKSPLASKEASRP 3372

Qy    509 -SGLKKLSGKKQKGKRGGGGDE--EPGEYQHIHTESPESADEQKGESSASSPEEPEETTC 565
Db   3373 TSVAESVKDEAEKSKEESSRDSVAEKSPLASKEASRPASVAESVQDEAEKSKEESRRESV 3432

Qy    566 LEKGPL---EAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRP--------- 613
Db   3433 AEKSPLASKEASRPASVAESVKDDAEKSKEESRRESVAEKSPLASKEASRPASVAESVKD 3492

Qy    614 -SESDKEE---ELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEAL 669
Db   3493 EAEKSKEESRRESVAEKSPLPSKEASRPTSVAESVKDEAEKSK-EESRRE---SVAEKSS 3548

Qy    670 ICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAV-PASTQE--QDQAQ 726
Db   3549 LASKEASRPASVAESVKDEA--EKSKEESRRESVAEKSPLASKEASRPASVAESVKDEAE 3606

Qy    727 GS-----------SSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSV--- 772
Db   3607 KSKEVSRRESVAEKSPLPSKEASRPTSVA--ESVKD-EADKSKEESRRESGAEKSPLASM 3663

Qy    773 ---------EQLSTEIEPSREES---WVSIKKFIPGRRKKR----ADGKQEQA--TVEDS 814
Db   3664 EASRPTSVAESVKDETEKSKEESRRESVTEKSPLPSKEASRPTSVAESVKDEAEKSKEES 3723

Qy    815 GPVEINEDDP------NVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTL 868
Db   3724 RRESVAEKSPLASKESSRPASV-------AESIKDEAEGTKQ-----------ESRRESM 3765

Qy    869 VHTVSVAVIDGTR-AVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA---- 923
Db   3766 PESGKAESIKGDQSSLASKETSRPDSVVESVKDETEKPEGSAIDKSQVASRPESVAVSAK 3825

Qy    924 -EETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMV 982
Db   3826 DEKSPL--HSRPE----------SVADKSPDA--SKEASRSLSVAE-TASSPIEEGP--- 3867

Qy    983 SAVSQLTDSPDTTEEA--------TPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQ 1034
Db   3868 RSIADLSLPLNLTGEAKGKLPTLSSPIDVAEGDFLEVKAESSPRPAVLSKPAEFSQPDTG 3927

Qy   1035 VPATQTVQRTGSKALEKVEEVEEDSE----VLASEKEKDVMPKGPVQEAGAEHLAQGSET 1090
Db   3928 HTASTPVDE-ASPVLEEIEVVEQHTTSGVGATGATAETDLL---DLTETKSETVTKQSET 3983

Qy   1091 G---------QATPESLE------VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLT 1135
Db   3984 TLFETLTSKVESKVEVLESSVKQVEEKVQTSVKQAETTVTDSLEQLTK-----KSSEQLT 4038

Qy   1136 ------DSETNGSTPLADS--DTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKE 1187
Db   4039 EIKSVLDTNISNVTNLFSTAVETIEKKVQDVT-EKVIEKATEHVSEHVTTTGESSTETSQ 4097

Qy   1188 EPSTL-------------------------PNNVPA-------QEEHGEEPGRDV---- 1210
Db   4098 EKSSLDLGTFSELRETHITTVGSPEFTVTICERDEPVLHDIKEEDEEHRFSPPSDVDKAA 4157
```

```
Qy   1211 ---------LEPTQQELTAAAVPVLAKTEVGQEGEVDWLDG-------EKVKEEQEVFVH 1254
Db   4158 IIPPQPMRPLSPREEEV-AKIVADVAKVLKSDKDITDIIPDFDERQLEEKLKSTADTEEE 4216

Qy   1255 SGPNS--QKAADVTYDSEVMGVAGCQEKESTEVQS--LSLEEGEMETDVEKEK------R 1304
Db   4217 SDKSTRDEKSLEISVKVEIE-----SEKSSPDQKSGPISIEEKDKIEQSEKAQLRQGILA 4271

Qy   1305 ETKPEQVSEEGE------QETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQ 1358
Db   4272 SSRPESVASQPESVPSPSQSAASHEHK----------EVELSESHKAEKS-----SRPES 4316

Qy   1359 DKAGCIEVQVQSLDTTVTQTA-------------EAVEKVIETVVISETGESPECVGAHL 1405
Db   4317 -----VASQVSEKDMKTSRPASSTSQFSTKEGDEETTESLLHSLTTTETVETKQMEEKSS 4371

Qy   1406 LPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQRE- 1464
Db   4372 FESVSTSVTKS--TVLSSQSTVQLREESTSESL--------SSSLKVE-----DSSRRES 4416

Qy   1465 -RSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFARTE- 1522
Db   4417 LSSLLAEKGGIATNTSLKEDTS-ASASQLEELLVQSEECSSESIVSEIQTSIAQKSNKEI 4475

Qy   1523 -TAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAWA 1579
Db   4476 KDARET------------KVTSQFTTTTSSATKDDSLKETVAEFLATEKIVSAKEAFS 4521
```

RESULT 5
QRMSP1
microtubule-associated protein MAP1B - mouse
N;Alternate names: microtubule-associated protein MAP1(X); microtubule-associated protei
C;Species: Mus musculus (house mouse)
C;Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 01-Sep-2000
C;Accession: S07549; S44387; A33645
R;Noble, M.; Lewis, S.A.; Cowan, N.J.
J. Cell Biol. 109, 3367-3376, 1989
A;Title: The microtubule binding domain of microtubule-associated protein MAP1B contains
A;Reference number: A33645; MUID:90094539; PMID:2480963
A;Accession: S07549
A;Molecule type: mRNA
A;Residues: 1-2464 <NOB>
A;Cross-references: EMBL:X51396; NID:g52999; PIDN:CAA35761.1; PID:g53000
R;Sanchez, C.; Padilla, R.; Paciucci, R.; Zabala, J.C.; Avila, J.
Arch. Biochem. Biophys. 310, 428-432, 1994
A;Title: Binding of heat-shock protein 70 (hsp70) to tubulin.
A;Reference number: S44387; MUID:94234720; PMID:8179328
A;Accession: S44387
A;Status: preliminary
A;Molecule type: protein
A;Residues: 653-663,'IC' <SAN>
C;Superfamily: microtubule-associated protein MAP1B
C;Keywords: microtubule binding; phosphoprotein; tandem repeat
F;589-786/Domain: microtubule binding #status experimental <MTB>
F;589-592,639-642,649-652,655-658,660-663,668-671,674-677,679-682,683-686,687-690,691-69
R-K-E/D-X)
F;1861-2064/Region: 17-residue repeats
F;91,116,351,888,1124,1153,1168,1208,1662,1877,1918,2003,2030,2054,2083/Binding site: ph
F;147,969,1336,1562,1563,1702,1708,1990,2057,2063,2419/Binding site: phosphate (Thr) (co
F;1953/Binding site: phosphate (Tyr) (covalent) #status predicted

```
  Query Match            6.2%;  Score 501;  DB 1;  Length 2464;
  Best Local Similarity  21.1%;  Pred. No. 1.5e-10;
  Matches  387;  Conservative  266;  Mismatches  648;  Indels  532;  Gaps   86;

Qy     45 ADPATKL-----PQKNGQLSSVNGVAEQGDVH------VQEENQEGQEEEVVDEDVGQRE 93
Db    478 ANPAEKIIRVLFPGNSTQYNILEGLEKLKHLDFLKQPLATQKDLTGQVPTPPVKQVKLKQ 537

Qy     94 SEDVREKDRVEEMAANSTAVEDITKDGQEETSEI-----IEQIPASENNVEEMVQPAESQ 148
Db    538 RADSRES---LKPATKPVASKSVRKESKEETPEVTKTSQVEKTPKVESKE---------- 584

Qy    149 ANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETA 208
Db    585 ---------------KVLVKKDKPVKTESKPSVT---------EKEVSSKEEQSP-VKAE 619

Qy    209 VGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKS 268
Db    620 VAEKQATESKPKVTKDK---VVKKE-------IKTKLEEKKEEKPKKEVVKKEDKTPLKK 669

Qy    269 PESP-SSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEE 327
Db    670 DEKPRKEEVKKEIKKEIKK-------EERKELKKEVKKETPLKDAKKEVKKEEKK--EVK 720

Qy    328 KEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDE-- 385
Db    721 KEEKEPKKEIKKISKDIKKSTPQSDTKKPS-ALKPKVAKKEESTKKE--PLAAGKLKDKG 777

Qy    386 -----KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVP 440
Db    778 KVKVIKKEGKTTEAAATAVGTAATTAAVVAAAGIAASGPVKELEAER---SLMSSPEDLT 834

Qy    441 QEAEPAEELMKSREMCVSGGDHTQLTDLSPEEK---TLPKHPEGIVSEVEMLSSQERIKV 497
Db    835 KDFEE----LKAEEIDVAKDIKPQLELIEDEEKLKETQPGEAYVIQKETEVSKGSAE--- 887

Qy    498 QGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSP 557
Db    888 --SP------DEGITTTEGE-------GECEQTPEELEPVEKQGVD--DIEKFEDEGAGF 930

Qy    558 EEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESD 617
Db    931 EESSET-----GDYEEKAETEEAEEPEEDGEDNASG-----SASK--------HSPTEDD 972

Qy    618 KEEELE-----KVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICV 672
Db    973 ESAKAEADVHLKEKRESVVSGDDRAEEDMDDVLEKGEAEQSEEEGEEED----------- 1021

Qy    673 GSSKKRARKASSSDDEG-GPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSS-- 729
Db   1022 --------KAEDAREEGYEPDKTEAEDYVMAVADKAAEAGV------TEEQYGYLGTSAK 1067

Qy    730 -------SPEPAGS--------PSEGEGVSTWESFK-----RLVTPRKKSKSKLEEKAED 769
Db   1068 QPGIQSPSREPASSIHDETLPGGSESEATASDEENREDQPEEFTATSGYTQSTIEISSEP 1127

Qy    770 SSVEQLST-----------EIEPSREESWVSIKKF-------------------IPGRR 798
Db   1128 TPMDEMSTPRDVMSDETNNEETESPSQEFVNITKYESSLYSQEYSKPAVASFNGLSEGSK 1187

Qy    799 KKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAV 858
Db   1188 TDATDGKDYNASASTISPPSSMEEDKFSKSA--LRDAYCSEEKELKASAELDIKD----- 1240

Qy    859 YVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHT-AGE-----AMPP 912
Db   1241 -VSDE-------RLSPA-------------KSPS-LSPSPPSPIEKTPLGERSVNFSLTP 1278

Qy    913 VE-EVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTAT-----------E 959
Db   1279 NEIKVSAEGEARSVSPGVTQAVVEEHCASPEEKTLEVVSPSQSVTGSAGHTPYYQSPTDE 1338

Qy    960 TSEALRTE-----------EVTEASGAEETT-------------------------- 979
Db   1339 KSSHLPTEVSENAQAVPVSFEFSEAKDENERASLSPMDEPVPDSESPVEKVLSPLRSPPL 1398

Qy    980 --------DMVSAVSQL----TDSP--------DTTEEATPVQEVES-GVLDTEEEERQT 1018
Db   1399 LGSESPYEDFLSADSKVLGRRSESPFEGKNGKQGFPDRESPVSDLTSTGLYQDKQEEKST 1458

Qy   1019 QAILQAVADKVKEE-SQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQ 1077
Db   1459 GFI------PIKEDFGPEKKTSDVETMSSQSALALDERKLGGDV--SPTQIDVSQFGSFK 1510
```

```
Qy   1078 EAGAEHLAQGSETGQ-ATPESLEVPEVT-ADVDHVATCQVIKLQQLMEQAVAPESSETLT 1135
Db   1511 EDTKMSISEGTVSDKSATPVDEGVAEDTYSHMEGVASV-----------STASVATSSFP 1559

Qy   1136 DSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEE------- 1188
Db   1560 EPTTDDVSPSLHAEV--GSPHSTEVD--DSLSVSVVQTPTTFQETEMSPSKEECPRPMSI 1615

Qy   1189 ------PSTLPNNVPAQEEHGEEPGRDV-----------------LEPTQQELTAAAVPV 1225
Db   1616 SPPDFSPKTAKSRTPVQDHRSEQSSMSIEFGQESPEHSFAMDFSRQSPDHPTLGAS---V 1672

Qy   1226 LAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYD---SEVMGVAGCQEKES 1282
Db   1673 LHITENGPT-EVDY---SPCDIQDSSLSHKIPPTEEPS-YTQDNDLSELISVSQVEASPS 1727

Qy   1283 T-----------------------------------EVQSLSLEEGEMETDVEK-EKRE 1305
Db   1728 TSSAHTPSQIASPLQEDTLSDVVPPREMSLYASLASEKVQSLEGEKLSPKSDISPLTPRE 1787

Qy   1306 TKP-----EQVSEEGEQETAAPEHEGTYGKPV-----------------------LTLD 1336
Db   1788 SSPLYSPGFSDSTSAAKETAA-AHQASSSPPIDAATAEPYGFRSSMLFDTMQHHLALNRD 1846

Qy   1337 MPSS-----ERGKALG----------SLGGSPSLPDQDKAGCIEVQVQSLDT--TVTQTA 1379
Db   1847 LTTSSVEKDSGGKTPGDFNYAYQKPENAAGSPDEEDYD----YESQEKTIRTHDVVRYYY 1902

Qy   1380 EAVEKVIETVVISETGESPECVGAHLLPAEKSSAT--GGHWTLQHAEDTVPLGPESQAES 1437
Db   1903 EKTERTIKSP--CDSGYSYETI-------EKTTKTPEDGGYTCEITEKTTRT-PEEGGYS 1952

Qy   1438 IPIIVTPAPESTLHPDLQGEISASQRERSEE---------EDKPDAGPD-ADGKESTAIE 1487
Db   1953 YEI----SEKTTRTPEVSG-YTYEKTERSRRLLDDISNGYDDTEDGGHTLGDCSYSYETT 2007

Qy   1488 KVLKAEPEILELESKSNKIVLNVIQTAVDQFARTETAPETHAY-----DSQTQVPACRLD 1542
Db   2008 EKITSFP---ESESYS-----------YETSTKTTRSPDTSAYCYETMEKITKTPQASTY 2053

Qy   1543 SRE-PNRCWTKMKDAKMKHPVPQPREDLQVLTV 1574
Db   2054 SYETSDRCYTTEK----KSP-SEARQDVDLCLV 2081
```

RESULT 6
T34418
hypothetical protein F12F3.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34418
R;Fulton, B.; Wohldmann, P.
submitted to the EMBL Data Library, July 1998
A;Description: The sequence of C. elegans cosmid F12F3.
A;Reference number: Z21521
A;Accession: T34418
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-3488 <FUL>
A;Cross-references: EMBL:U80022; PIDN:AAC25885.1; GSPDB:GN00023; CESP:F12F3.3
A;Experimental source: strain Bristol N2; clone F12F3
C;Genetics:
A;Gene: CESP:F12F3.3
A;Map position: 5
A;Introns: 281/3; 332/1; 562/3; 600/3; 1866/3; 1944/3; 3393/1

```
  Query Match             6.2%;  Score 497.5;  DB 2;  Length 3488;
  Best Local Similarity   20.1%;  Pred. No. 3.1e-10;
  Matches  330;  Conservative  284;  Mismatches  568;  Indels  457;  Gaps   67;

Qy     80 QEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNV- 138
Db    269 RDGKLIDDGSGDR--MEVRHED---EMRKHWLILKDICKD--EEAEYACQAI-----NVA 316

Qy    139 -------EEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKN-EKSDTVQLLTVKKDEGEG 190
Db    317 GEAWCFSDVVVHMSEESRDD------------------DKSVDEVDDSTVLEEKKDDGD- 357

Qy    191 AEASVGAGDHQEPSV-ETAVGESASKESELKQSTEKQEGTLKQ--EQSSTEIPLQAESDQ 247
Db    358 --------DKSKPKTKKKIIKKKETPESEQVTAAEPEQQKISEVDVQSVAETEVGAKKKP 409

Qy    248 AAEE-----EAKDEGEEKQEKEPTKSPESPSS---PVNSETTSSFKKFFTHGWAGWRKKT 299
Db    410 DAEKPTDLSKAKKDSKSKKSDEPEASTEEKSTTEKPTNDKTS---------------KKS 454

Qy    300 SFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELP 359
Db    455 AEKKTVKPKKEVT--GKPLEAKKPVEDKKDASQPSS-SKESSPPTDGKKKKQIPKALFIP 511

Qy    360 --LEDQVGD----------------LEASSEEK---CAPLATEVFDEKMEAHQEVVAEVH 398
Db    512 DEISSRFGDPSTMHSETNITTTIRGREGSADAKTPLVEPLSASVSMKVFTLVESAKEKAE 571

Qy    399 VSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQ-EVPQEAEPAEELMKSREMCV 457
Db    572 FSFKRRSETPDDKSRKKEG----------LPPAKKSEKKDEVTAEKQSTEALIESKKKEV 621

Qy    458 SGGDHTQLTDLSPEEKT------LPKHPEG-----IVSEVEMLSSQERIKVQGSPLKKLF 506
Db    622 ---DESKISEQQPSDKNKSEVVGVPEKAAGPETKKDVSEIEEV----------PKKKTI 667

Qy    507 SSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
Db    668 KKKTEKSDSSISQKSNVLKPADDDKSKSDDVTDKSKKTTEDQTKVATDSKLEKAADTT-- 725

Qy    567 EKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK 626
Db    726 KQIETETVVDDKSKKKVL---KKKTEKSDSFISQKSETPPVVEPTKPAESEAQKIAEVNK 782

Qy    627 SATLSSTDSTVS-EMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSS 685
Db    783 AKKQKEVDDNLKREAEVAAKKIADEKLKIEAEANIKKTAEVEA----------AKKQKEK 832

Qy    686 DDE-------GGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPS 738
Db    833 DEQLKLETEVVSKKSAAEKLELEKQAQIKKAAEADAVKKQKELNEKNKLEAAKKSAADKL 892

Qy    739 EGEGVSTWESFK---RLVTPRKKSKSKLEEK-----AEDSSVEQLSTEIEPSREESWVSI 790
Db    893 KLEEESAAKSKKVSEESVKFGEEKKTKAGEKTVQVESEPTSKKTIDTKDVGATEPADETP 952

Qy    791 KKFIPGRRKKRADGKQEQATVEDSGPV--EINEDDPNVPAVVPLSEYNAV-EREKMEAQG 847
Db    953 KKKIIKKKTEKSDSSISQKSATDSEKVSKQKEQDEPTKPAV---SETQMVTEADKSKKQK 1009

Qy    848 NTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAG 907
Db   1010 ETD-----------EKLK-----------LDAEIAAKTKQEAD------------EKSKL 1035

Qy    908 EAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTE 967
Db   1036 DAQEKIKKVSEDDAARKE-----------KELND-----KLKLESEIATKKASADKLKLE 1079

Qy    968 EVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTE-----------EEER 1016
Db   1080 EQAQAKKAAE----VEAAKKQKEKDEQLK------------LDTEAASKKAAAEKLELEK 1123

Qy   1017 QTQAILQAVAD------------------------KVKEESQVPATQTVQ-------RT 1044
Db   1124 QAQIKKAAGADAVKKQKELDEKNKLEANKKSAAGKLKIEEESAAKSKQTVEEQAKLDAQT 1183

Qy   1045 GSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQ-------EAGAEHLAQGSETGQATPES 1097
Db   1184 KAKTAEKQTKLEKDEKSTKESESKETVDEKPKKKVLKKKTEKSDSSISQKSETSKTVVES 1243
```

```
Qy   1098 LEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSET--LTDS-----ETNGSTPLADSDT 1150
                                              |  |||   : |:      ||:    |     |
Db   1244 ---------------------------AGPSESETQKVADAARKQKETDEKQKLEAEIT 1275

Qy   1151 ADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDV 1210
           |  :  ::          ||  |   :  :  | |||  |||:    | :     |     ::   :
Db   1276 AKKSADEK------SKLEAESKLKKAAEVEAAKKQKEKDEQLKLDT---EAASKKAAAEK 1326

Qy   1211 LEPTQQELTAAAVPVLA---------KTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQK 1261
           || :|       |  | |           |  :   |       | ||:| |::            :|
Db   1327 LELEKQSHIKKAAEVDAVKKQKELEEKQRLESEAATKKADAEKLKLEEQ--------KKK 1378

Qy   1262 AADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAA 1321
           ||::         :: :   ||| :  |  ||     :|  :     : :|  : ||| :|    |          |
Db   1379 AAEIA----LIEIQKEQEKLAQE-QSRLEDEAKKSAEKQKLESETKSKQTEE-------A 1426

Qy   1322 PEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQ---- 1377
           |:                                              |: ::  |      :: :  :    |::::|
Db   1427 PKE---------------------------SVDEKPKKKVLKKKTEKSDSSISQKSKS 1457

Qy   1378 ---TAEAVE------KVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVP 1428
              | :| |          ::|    :  :  :||:               | :|||        |     |:
Db   1458 AKSTVDAAETLESDFNLVEKKTVQKVEQSPD---------ESTSATIKRDPAQKTEEISK 1508

Qy   1429 LGPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDA-GPDA--------- 1478
                : ::             | || :             | :  :|      ::    | |:     ||
Db   1509 QDDGDEKKTTTDGKPPKPEDS-------EATPKKRVVKKKTQKSDSVASDASLADVSKLS 1561

Qy   1479 DGKESTAIEKVLK-----------------AEPEILELES-KSNKIVLNVIQTAVDQFA 1519
           |    |      :||||                                 :|| :|: :  |: :::|:            ::|:
Db   1562 DDVEEKPKKKVLKKKTEKSDSVISETSSVDTIKPESVEIPTEKAEQMILH------NRFS 1615

Qy   1520 RTETAPET---HAYDSQTQ 1535
            |::| |:   :|:    |:
Db   1616 -TDSAVESEPKNAHKDDTE 1633
```

RESULT 7
S37431
ankyrin 2, neuronal long splice form - human
N;Alternate names: ankyrin B, 440K splice form; ankyrin-B; brain ankyrin; non-erythroid
N;Contains: ankyrin 2, short form
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 13-Aug-1999
C;Accession: S37431; A39643; B39643; A40334; A49462; S14533; S14569
R;Chan, W.
submitted to the EMBL Data Library, September 1993
A;Reference number: S37431
A;Accession: S37431
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-3924 <CHA>
A;Cross-references: EMBL:Z26634; NID:g406287; PIDN:CAA81387.1; PID:g406288
R;Otto, E.; Kunimoto, M.; McLaughlin, T.; Bennett, V.
J. Cell Biol. 114, 241-253, 1991
A;Title: Isolation and characterization of cDNAs encoding human brain ankyrins reveal a
A;Reference number: A39643; MUID:91302466; PMID:1830053
A;Accession: A39643
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2077 <OT1>
A;Cross-references: GB:X56957
A;Accession: B39643
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1443,3585-3924 <OTT>
A;Cross-references: EMBL:X56958
R;Tse, W.T.; Menninger, J.C.; Yang-Feng, T.L.; Francke, U.; Sahr, K.E.; Lux, S.E.; Ward,
Genomics 10, 858-866, 1991
A;Title: Isolation and chromosomal localization of a novel nonerythroid ankyrin gene.
A;Reference number: A40334; MUID:92009921; PMID:1833308
A;Accession: A40334
A;Molecule type: DNA
A;Residues: 463-474,'PE',477-495 <TSE>
A;Cross-references: GB:M37123; NID:g178647; PIDN:AAA62828.1; PID:g178648
R;Chan, W.; Kordeli, E.; Bennett, V.
J. Cell Biol. 123, 1463-1473, 1993
A;Title: 440-kD ankyrinB: structure of the major developmentally regulated domain and sr
A;Reference number: A49462; MUID:94075409; PMID:8253844
A;Accession: A49462
A;Status: preliminary; nucleic acid sequence not shown
A;Molecule type: mRNA
A;Residues: 1-3924 <RES>
A;Cross-references: EMBL:Z26634; NID:g406287; PIDN:CAA81387.1; PID:g406288
C;Genetics:
A;Gene: GDB:ANK2
A;Cross-references: GDB:127607; OMIM:106410
A;Map position: 4q25-4q27
C;Superfamily: ankyrin; ankyrin repeat homology
C;Keywords: alternative splicing
F;2-3924/Product: ankyrin 2, long form #status predicted <MAT>
F;2-1443,3585-3924/Product: ankyrin 2, short form #status predicted <MA2>
F;63-95/Domain: ankyrin repeat homology <AN01>
F;96-128/Domain: ankyrin repeat homology <AN02>
F;129-161/Domain: ankyrin repeat homology <AN03>
F;162-190/Domain: ankyrin repeat homology <AN04>
F;191-223/Domain: ankyrin repeat homology <AN05>
F;232-264/Domain: ankyrin repeat homology <AN06>
F;265-297/Domain: ankyrin repeat homology <AN07>
F;298-330/Domain: ankyrin repeat homology <AN08>
F;331-363/Domain: ankyrin repeat homology <AN09>
F;364-396/Domain: ankyrin repeat homology <AN10>
F;397-429/Domain: ankyrin repeat homology <AN11>
F;430-462/Domain: ankyrin repeat homology <AN12>
F;463-495/Domain: ankyrin repeat homology <AN13>
F;496-528/Domain: ankyrin repeat homology <AN14>
F;529-561/Domain: ankyrin repeat homology <AN15>
F;562-594/Domain: ankyrin repeat homology <AN16>
F;595-627/Domain: ankyrin repeat homology <AN17>
F;628-660/Domain: ankyrin repeat homology <AN18>
F;661-693/Domain: ankyrin repeat homology <AN19>
F;694-726/Domain: ankyrin repeat homology <AN20>
F;727-759/Domain: ankyrin repeat homology <AN21>
F;760-792/Domain: ankyrin repeat homology <AN22>
F;793-825/Domain: ankyrin repeat homology <AN23>

```
  Query Match             6.1%;  Score 493;  DB 2;  Length 3924;
  Best Local Similarity   20.4%;  Pred. No. 5.1e-10;
  Matches  422;  Conservative  270;  Mismatches  680;  Indels  698;  Gaps   99;

Qy     68 GDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSE- 126
           | : |:|  :  :||      :: :|  ||:    ::| |: :  |     |       |    ||
Db   1497 GSIKVKELVKAAEEEPGEPFEIVERVKEDL---EKVNEILRSGTCTRD-ESSVQSSRSER 1552

Qy    127 -IIEQ--IPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDK----NEKSDTV- 178
            ::|:  :  |:  :||  |  |   :   :   :|        |   |:||       |  :| :
Db   1553 GLVEEEWVIVSDEEIEEARQKAPLEITEYPCVEVRIDKEIKGKVEKDSTGLVNYLTDDLN 1612

Qy    179 --------QLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKES-----ELKQSTEK 225
                   || ||:    |:  ||                |||  |: ||        :  |||:|
Db   1613 TCVPLPKEQLQTVQDKAGKKCEA-------------LAVGRSSEKEGKDIPPDETQSTQK 1659

Qy    226 Q-----------EGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPE---- 270
           |                 ||::|   |  ||| :::|   |   |    ||     ::|   :||
Db   1660 QHKPSLGIKKPVRRKLKEKQKQKEEGLQASAEKA--ELKKGSSEESLGEDPGLAPEPLPT 1717

Qy    271 -SPSSPVNSET-TSSFKKFFTHGWAGWRKKTSFKKSKEDD------LETAEKRKEQEAEK 322
              :||:  ||     | |                  |   : :|  ||:        |     | ||     :|
Db   1718 VKATSPLIEETPIGSIKD----------KVKALQKRVEDEQKGRSKLPIRVKGKEDVPKK 1767

Qy    323 VDEEEKEKTEPASEEQ-----EPAEDTDQARLSADYEKVE------------------- 357
                          |: : :         |:   |: ::        |  |
Db   1768 TTHRPHPAASPSLKSERHAPGSPSPKTERHSTLSSSAKTERHPPVSPSSKTEKHSPVSPS 1827

Qy    358 LPLEDQVGDLEASSEEKCAPLA--------TEVFDEKMEAHQEVV-------------- 394
```

```
Db  1828 AKTERHSPASSSSKTEKHSPVSPSTKTERHSPVSSTKTERHPPVSPSGKTDKRPPVSPSG 1887

Qy   395 -AEVH------------VSTVEKTEEEQ--GGGGEAEGGVVVEGTG--ESLPPEKLAEP 436

Db  1888 RTEKHPPVSPGRTEKRLPVSPSGRTDKHQPVSTAGKTEKHLPVSPSGKTEKQPPVSPTSK 1947

Qy   437 QEVPQEAEPAEELMKSREMCVSGGD----------HTQLTDLSPEEK-----TLPKHP-- 479

Db  1948 TERIEETMSVRELMKAFQ---SGQDPSKHKTGLFEHKSAKQKQPQEKGKVRVEKEKGPIL 2004

Qy   480 ---EGIVSEVEMLSSQERIKVQGSPLKK---LFSSSGLKKLSGKKQKGKRGGGGDEE--- 530

Db  2005 TQREAQKTENQTIKRGQRLPVTGTAESKRGVRVSSIGVKK--------EDAAGGKEKVLS 2056

Qy   531 ----------PGEYQHIHTESPES--ADEQKGESSASSPEE------------------ 559

Db  2057 HKIPEPVQSVPEEESHRESEVPKEKMADEQGDMDLQISPDRKTSTDFSEVIKQELEDNDK 2116

Qy   560 ------PEETTCLEKGPLEAPQ-----------------------------DGEAE--- 580

Db  2117 YQQFRLSEET---EKAQLHLDQVLTSPFNTTFPLDYMKDEFLPALSLQSGALDGSSESLK 2173

Qy   581 -------------EGT---TSDGEKKREGI--TPWASFKKM-VTPKK---RVRRPSESDK 618

Db  2174 NEGVAGSPCGSLMEGTPQISSEESYKHEGLAETPETSPESLSFSPKKSEEQTGETKESTK 2233

Qy   619 EEELEKVKS------------------ATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Db  2234 TETTTEIRSEKEHPTTKDITGGSEERGATVTEDSETSTESFQKEATLG-SPKDTSPKRQD 2292

Qy   661 DTSVSWEALICVGSSKKRARKAS-------SSDDEGGPRTLGGDSHRAEEASKDKEAGTD 713

Db  2293 D---------CTGSCSVALAKETPTGLTEEAACDE-GQRTFGSSAHKTQTDSEAQE---- 2338

Qy   714 AVPASTQEQDQAQGSSSPEPA-----GSPSEGEGVSTWESFKRLVTPRKKS--------- 759

Db  2339 ----STATSDETKALPLPEASVKTDTGTESKPQGVIRSPQGLELALPSRDSEVLSAVADD 2394

Qy   760 ------KSKLEEK--AEDSSVEQLSTEIEPS--REESWVSIKKFIPGRRKKRA------- 802

Db  2395 SLAVSHKDSLEASPVLEDNSSHKTPDSLEPSPLKESPCRDSLESSPVEPKMKAGIFPSHF 2454

Qy   803 --------------------------------DGKQEQATVEDSGPVEINEDDPN----- 825

Db  2455 PLPAAVAKTELLTEVASVRSRLLRDPDGSAEDDSLEQTSLMESSGKSPLSPDTPSSEEVS 2514

Qy   826 -------------VPAVV--------------------------PLSEYNAVERE--- 841

Db  2515 YEVTPKTTDVSTPKPAVIHECAEEDDSENGEKKRFTPEEEMFKMVTKIKMFDELEQEAKQ 2574

Qy   842 --------KMEAQGNTELPQLLGAVYVSEELSKTLVHT--------VSVAVIDGTRAVTS 885

Db  2575 KRDYKKEPKQEESSSSSDPDADCSVDVDEP-----KHTGSGEDESGVPVLVTSESRKVSS 2629

Qy   886 VEERSPSW----------------ISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVL 929

Db  2630 SSESEPELAQLKKGADSGLLPEPVIRVQPPSPLPSSMDSNSSP-EEVQFQPVVSKQ---- 2684

Qy   930 TQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTE----------------EVTEAS 973

Db  2685 -YTFKMNEDTQEEPGKSEEEKDSESHLA-EDRHAVSTEAEDRSYDKLNRDTDQPKICDGH 2742

Qy   974 GAEETTDMVSA--VSQLTDSP-----------------------DTTEEATPVQEVESG 1007

Db  2743 GCEAMSPSSSARPVSSGLQSPTGDDVDEQPVIYKESLALQGTHEKDTEGEELDVSRAESP 2802

Qy  1008 VLDTEEEERQTQAILQ--AVADKVKEESQVPATQTVQRTG-SKALEKVEEVEED------ 1058

Db  2803 QADCPSESFSSSSSLPHCLVSEGKELDEDISATSSIQKTEVTKTDETFENLPKDCPSQDS 2862

Qy  1059 ---------------SEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATP-------E 1096

Db  2863 SITTQTDRFSMDVPVSDLAENDEIYDPQITSPYENVPSQSFFSSEESKTQTDANHTTSFH 2922

Qy  1097 SLEVPEV--TADVDHVATC----------------QVIKLQQLMEQAVAPESSETLTDSE 1138

Db  2923 SSEVYSVTITSPVEDVVVASSSGTVLSKESNFEGQDIKMESQLESTL----WEMQSDSV 2978

Qy  1139 TNGSTP-LADSDTADGTQQDETIDSQ---DSKATAAVR------QSQVTEEE----AATA 1184

Db  2979 SSSFEPTMSATTTVVGEQISKVIITKTDVDSDSWSEIREDDEAFEARVKEEEQKIFGLMV 3038

Qy  1185 QKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQEL-TAAAVPVLAKT----EVGQEGEVDW 1239

Db  3039 DRQSQGTTPDTTPAR------------TPTEEGTPTSEQNPFLFQEGKLFEMTRSGAID- 3085

Qy  1240 LDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVM-GVAGCQEKE-STEVQSLSLEEGEMET 1297

Db  3086 MTKRSYADESFHFFQIG---QESREETLSEDVKEGATGADPLPLETSAESLALSESKETV 3142

Qy  1298 DVEKEKRETKPEQVSEEGEQ----------------ETAAPEHEGTYGKPVLTLDMPSSE 1341

Db  3143 D---DEADLLPDSVSEEVEEIPASDAQLNSQMGISASTETPTKEAV---SVGTKDLPTVQ 3196

Qy  1342 RGKALGSLGG--SPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETG-ESP 1398

Db  3197 TGD-IPPLSGVKQISCPDSSEPA---VQVQLDFSTLTRS-----------VYSDRGDDSP 3241

Qy  1399 ECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGP-------ESQAESIPI----IVTPAPE 1447

Db  3242 DS-----SPEEQKSVI-----------EIPTAPMENVPFTESKSK-IPVRTMPTSTPAPP 3284

Qy  1448 STLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL-KAEPEILELESKSNKI 1506

Db  3285 SA---EYESSVSEDFLSSVDEENKAD---EAKPKSKLPVKVPLQRVEQQLSDLDTSVQKT 3338

Qy  1507 VL----NVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPV 1562

Db  3339 VAPQGQDMASIAPDNRSKSES--DASSLDSKTKCPV-------KTRSYTETETESRER-- 3387

Qy  1563 PQPREDLQVLTVLEAWAQPRKCLPRLQLKA 1592

Db  3388 ---AEELE-LESEEGATRPKILTSRLPVKS 3413
```

RESULT 8
T29757
protein UNC-89 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 03-Dec-1999
C;Accession: T29757
R;Du, Z.; Le, T.T.; Wilson, R.
submitted to the EMBL Data Library, May 1997
A;Description: The sequence of C. elegans cosmid C09D1.
A;Reference number: Z20679
A;Accession: T29757
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-6642 <DUZ>
A;Cross-references: EMBL:AF003131; PIDN:AAB54132.1; GSPDB:GN00019; CESP:unc-89
A;Experimental source: strain Bristol N2; clone C09D1
C;Genetics:
A;Gene: CESP:unc-89
A;Map position: 1
A;Introns: 17/2; 108/3; 154/2; 211/2; 265/3; 326/2; 352/3; 426/2; 454/1; 500/1; 537/1; 6
/3; 5917/1; 6027/1; 6061/3; 6153/2; 6515/1; 6552/3; 6609/1

  Query Match 6.1%; Score 489; DB 2; Length 6642;
  Best Local Similarity 21.8%; Pred. No. 1.3e-09;
  Matches 348; Conservative 234; Mismatches 670; Indels 346; Gaps 68;

```
Qy     7 TEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSSVNGVAE 66

Db  1132 TVQKKPFAPEFDLRPVSLTVEKGSEAVFSAHAFGIP--------LPTYEW---SVNG--- 1177
```

```
Qy     67 QGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKD-----------RVEEMAANSTAVED 115
Db   1178 -------RKVRDGQEGARVTRDESTVDGASILTIDTATYYSEVNHLTISVVAENTLGAEE 1230

Qy    116 ----ITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKF--TVKK 169
Db   1231 TGAQLTIEPKKE-SVVVEKQDLSSSEVQKEIAQQVKEASPEATTTITMETSLTSTKTTTM 1289

Qy    170 DKNEKSDTVQLLTVKKDEGEGAEASV----GAGDHQEPSVETAVGESASKESELKQSTEK 225
Db   1290 STTEVTSTVGGVTVETKESESESATTVIGGGSGGVTEGSISVSKIEVVSKTD---SQTDV 1346

Qy    226 QEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFK 285
Db   1347 REGTPKRRVSFAEEELPKEVIDSDRKKKKSPSPDKKEKSPEKTEEKPASPT--------- 1397

Qy    286 KFFTHGWAGWRKKT--SFKKSKE-DDLETAEKRKEQEAEKV---DEEEKEKTEPASEEQE 339
Db   1398 -----------KKTGEEVKSPKEKSPASPTKKEKSPAAEEVKSPTKKEKSPSSPTKKEKS 1446

Qy    340 PAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHV 399
Db   1447 PSSPT---KKTGDEVKEKSPPKSPT--KKEKSPEKPEDVKSPVKKEKSPDATNIVEVSSE 1501

Qy    400 STVEKTE-----EEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKSRE 454
Db   1502 TTIEKTETTMTTEMTHESEESRTSVKKEKT-----PEKVDEKPKSPTKKDKSPEKSITEE 1556

Qy    455 MCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERI-KVQGSPLKKLFSSSGLKK 513
Db   1557 I-----------KSPVKK--EKSPEKVEEKPASPTKKEKSPEKPASPTKK---SENEVK 1599

Qy    514 LSGKKQKGKRGGGDE--EPGEYQHIHTESPESAD---------EQKGESSAS------- 555
Db   1600 SPTKKEKSPEKSVVEELKSPKE------KSPEKADDKPKSPTKKEKSPEKSATEDVKSPT 1653

Qy    556 ----SPEEPEE----TTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPK 607
Db   1654 KKEKSPEKVEEKPTSPTKKESSPTKKTDD-EVKSPTKKEKSPQTVEEKPASPTKKEKSPE 1712

Qy    608 K----RVRRPSESDKEEELEKVKSAT---LSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
Db   1713 KSVVEEVKSPKEKSPEKAEEKPKSPTKKEKSPEKSAAEEVKSPTKKEKSPEKSAEEKPKS 1772

Qy    661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
Db   1773 PT------------KKESSPVKMADDE-------VKSPTKKEKSPEK---VEEKPASPT 1809

Qy    721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
Db   1810 KKEKTPEKSAAEELKSPTKKE--------KSPSSPTKKTGDESKEKSPE----------K 1851

Qy    781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAV--VPLSEYNAV 838
Db   1852 PEEKPKSPTPKKSPPGSPKK----KKSKSPEAEKPPAPKLTRDLKLQTVNKTDLAHFEVV 1907

Qy    839 EREKMEAQGNTELPQLLGA--VYVSEE--------LSKTLVHTVSVAVIDGTRAVTSVEE 888
Db   1908 VEHATECKWFLDGKEITTAQGVTVSKDDQFEFRCSIDTTMFGSGTVSVV-ASNAAGSVET 1966

Qy    889 RSPSWISASVTEPLEHTAGEAMPPVE----EVTEKDIIAEETPVLT-----QTLPEGKDA 939
Db   1967 KTELKVLETPKETKKPEFTDKLRDMEVTKGDTVQMDVIALHSPLYKWYQNGNLLEDGKNG 2026

Qy    940 HDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEAT 999
Db   2027 --VTIKNEENKSSLIIPNAQDSGKITVEASNEVGSSESS-------AQLTVNPPST---T 2074

Qy   1000 PVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDS 1059
Db   2075 PI--VVDGPKSVTIKETETAEFKATI-------SGFPA-PTVKWTIN------EKIVEES 2118

Qy   1060 EVLASEKEKDV----MPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQV 1115
Db   2119 RTITTIKTEDVYTLKISNAKIEQTGTVKVTAQNSAGQDSKQ--------ADLKVEPNVKA 2170

Qy   1116 IKLQQLMEQAVAPESSETLTDSETNGST------------PLADSDT------ADGTQQ- 1156
Db   2171 PKFKSQLTDKVADEGEPLRWNLELDGPSPGTEVSWLLNGQPLTKSDTVQVVDHGDGTYHV 2230

Qy   1157 --DETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPT 1214
Db   2231 TIAEAKPEMSGTLTAKAKNAAGECETSAKVTVNGGNKKPEFVQAPQNH---------ETT 2281

Qy   1215 QQELTAAAVPVLAKTEVGQEGEVDW-LDGEKVKEEQEV---FVH-SGPNSQKAADVTYDS 1269
Db   2282 LEESVKFSAIVTGK----PMPNVTWYLNNKKLIQSEEVKVKYVHETGKTS-----IRIQK 2332

Qy   1270 EVMGVAGC----QEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHE 1325
Db   2333 PLMEHNGTIRVEAENVSGKVQATAQLKVDKKTEVPKFTTNMDDRQV-KEGEDVKFTANVE 2391

Qy   1326 GTYGKPVLTLDMPSSERGKALGSLGGSP-------SLPDQDKAGCIEV------QVQSLD 1372
Db   2392 G-YPEP------------SVAWTLNGEPVSKHPNITVTDKDGEHTIEISAVTPEQAGELS 2438

Qy   1373 TTVTQTAEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPE 1432
Db   2439 CEATNPVGSKKRDVQ-LAVKKVGDAP----TFAKNLEDRLITEGELTLMDAKLN------ 2487

Qy   1433 SQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEED 1470
Db   2488 --------IVKPKPKITWLKDGVEITSDGHYKIVEEED 2517
```

RESULT 9
T15348
hypothetical protein B0350.1 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C;Accession: T15348
R;Gattung, S.
submitted to the EMBL Data Library, February 1996
A;Description: The sequence of C. elegans cosmid B0350.
A;Reference number: Z18332
A;Accession: T15348
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-5170 <GAT>
A;Cross-references: EMBL:U50071; NID:g1208871; PID:g1208877; PIDN:AAA93447.1; CESP:B035C
C;Genetics:
A;Gene: CESP:B0350.1
A;Introns: 48/1; 5039/3; 5116/3

```
  Query Match             6.0%;  Score 488;  DB 2;  Length 5170;
  Best Local Similarity   20.0%;  Pred. No. 1.1e-09;
  Matches  391;  Conservative  249;  Mismatches  678;  Indels  634;  Gaps   89;

Qy      5 SSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGD----PADADPATKLPQKNGQLSS 60
Db    818 AESERQVPESPVVSHQEYPHVVETTTTNVTSNIYDDEDNVPSSEDPATQHFQQ-----S 872

Qy     61 VNGVAEQGDVHVQEENQEGQEEEVVD------------------EDVGQRESEDVREKD 101
Db    873 ETSVHRSHPDSVEESDGEGLGSKVLGFAKKAGMVAGGVVAAPVALAAVGAKAAYDAFEKD 932

Qy    102 RVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFK------ 155
Db    933 -----------------DEDDETSHSPES-PVPEYQ-SEQYQDDSAQSSHTDFEHHMPES 973

Qy    156 --------KVFKFVGFKFTVKKDKNEK----SDTVQLLTVKK---DEGEGAEASVGAGDH 200
Db    974 PIEHEKETEEFDHSHPESPVLSEKEREHQVTSETTTTTVTREYNDEPEDEQ------EH 1027

Qy    201 QEPS---------------VETAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAES 245
```

```
Db   1028 QGPHSPAPSSHTEAEHPHIVETTTTTTVTREFQ----EEPEELEYKQEDNSRKSPSSHSQ 1083

Qy    246 DQAAEE---------EAKDEGEEKQEKEPTKSPESPSSPVNSE---------------- 279

Db   1084 ENLVTETTTTTTVTSEYYDEPEHFEEQETGKSSPAPSSHVESERQVLESPVASDPRHVME 1143

Qy    280 --TTSSFKKFFTHGWAGWRKKTSFKKSKE-DDLETAEKRKEQEAEKVDEEEKEKTEPASE 336

Db   1144 TTTTTTVTRQF-HDDESERSDSPNRDVVEAQSIHSSNIESHQQFSEKDDDSQRGPSPVKS 1202

Qy    337 EQEPAE------------------DTDQARLSADYEKVELPLE---------------- 361

Db   1203 EDEPVKHESYKFETSTTETREFDRPELDSERLS---EPAQSPMEVSPITTEENIVRPSSS 1259

Qy    362 -----DQVGDLEASSEEKCAPLAT-EVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGE- 414

Db   1260 VKSEYGSEGHVPSVIETKTTTTVTREFYDDDQDEHENQTQSEELRSSIPTEEEEHEGSHL 1319

Qy    415 ------------------------------AEGGVVVEGTGESLPPEKLAEPQEVPQ-- 441

Db   1320 FKETTTTTTVTREFYDEPENVEELQDPQFSPAPSSHVESGIHASESP--VAQQQEIPQTR 1377

Qy    442 ---EAEPAEELMKSREMCVSGGDHTQLTDLSP---EEKTLPKHPEGIVSEVEMLSSQERI 495

Db   1378 EFHEDSPAAQYFHEEEY-----ESHVLTEQAPLLTEQQHQPE--SGDESDGEGLGS---- 1426

Qy    496 KVQGSPLKKLFSSSG-------LKKLSGKKQKGKRGGGGDEEPGE------YQHIHTESP 542

Db   1427 KVLGFAKKAGMVAGGVVAAPVALAAVGAKAAYDALKKDDDEEDQEERESLLRQERSIDSP 1486

Qy    543 ESAD----EQKGESSASSPEEPE-----ETTCL---------EKGPLEAPQDGEAEEGTT 584

Db   1487 HASEQSQIEEEHERFEESPVPSEKHHVTETTTTTTVTREYFDEHEPL-VSQEIEGEKGSP 1545

Qy    585 SDGEK------------------KREGITPWASFKKM--VTPKKRVRRPSESDKEEELE 623

Db   1546 ADSEKSLPHVVETTTTTTVTREFDKNDSESPVPSEKEQEPTTVSREVYETAEGDEPEH-- 1603

Qy    624 KVKSATLSSTDSTVSEMQDEVKTVGEEQKP-EEPKRRVDTSVSWEALICVGSSKKRARKA 682

Db   1604 ---HYTETTTTTVTKEVIDDSQEMGDDDEPKQESPQVVETT---------TTTMTSREY 1650

Qy    683 SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAV-------PASTQEQDQAQGSSSPEPAG 735

Db   1651 DDNDDE--TRSEAGDSHITETTKT-----TTVVREFHGEQPEETEETDEVEE-------- 1695

Qy    736 SPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIP 795

Db   1696 -----------------LPPKIEEEDNVSEYSESST--SVSREVRPD-EPHIIETTTTTT 1735

Qy    796 GRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLL 855

Db   1736 VTREYHNEPEETYDDQKDAAPISFSQ------------EHQDDDSQASHDQHDRESP--- 1780

Qy    856 GAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEE 915

Db   1781 ------VESEKSVKHT--------TETTT----------TTTVTRQLYD---------DE 1807

Qy    916 VTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGA 975

Db   1808 ASE---IRGESPVATE-------------------EHEHVSSTKSDD----ESEQHVPSV 1841

Qy    976 EETTDMVSAVSQLTDSPD--TTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEES 1033

Db   1842 IETTTTTTVTREFYDDQDELQREDHTQSEERRSSI-PTEETEHEDSHLIK---------- 1890

Qy   1034 QVPATQTVQRTGSKALEKVEEVEEDSEVL--ASEKEKDV-MPKGPVQEAGAEHLAQGSET 1090

Db   1891 ETTTTTTVTREFYDEPENVEKLQDSQFSLSPSSHVESEIYVPESPV--AKQQEIPQTREF 1948

Qy   1091 GQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDT 1150

Db   1949 HEDSPAAQYFHE--DEYEHQVPTEQAPL--LTEQQHQPESGEE-SDGEGFGSKVLGFAKK 2003

Qy   1151 A----------------------DGTQQDETIDSQDSKATAAVRQSQVTEE-------- 1179

Db   2004 AGMVAGGVVAAPVALAAVGAKAAYDALKKEDDEEDQEE-----RESLIREERSFDSPHA 2057

Qy   1180 -EAATAQKEEPSTLPNNVPAQEEHGEE---PGRDVLEPTQQELTAAAVPVLAKTEVGQEG 1235

Db   2058 SEQSQIEKEHKRFEESPVPSEKHHDQSSALPQESVSQPIEKE----------SRTFNDES 2107

Qy   1236 EVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEM 1295

Db   2108 EFGVKSDHYTEDDQESLKSPKESGEAFSQFTSEKE-------QDRSDSPIHS-------Q 2153

Qy   1296 ETDVEKEKRETKPEQVSEEGEQETAAP--EHEGTYGKPVLTLDMPSSERGKALG-----S 1348

Db   2154 KEDISQFQNESSPEDVKSEQPHDEEKPDLERQGSYSSGY----SPKSPGGSITGLDEEKA 2209

Qy   1349 LGG---------------------------------SPSLPDQDKAGCIEVQVQSLDT 1373

Db   2210 LSGVQEPEDRPENFAESHEKTEATSDENLFESDKYAPASP-VPSEDSSN----RVIETTT 2264

Qy   1374 TVTQTAEAVEK------VIETVVISETGESPECVGAHLLPAEKS------SATGGHWTLQ 1421

Db   2265 TTTVTREHFEPEDDHSYVVESQEYSSSG-SP-------VPSEKSVDRVIETTTTTTVTRE 2316

Qy   1422 H--AEDTVPLGPES----QAESIPIIVTPAP-ESTLHPDLQGEISAS------------ 1461

Db   2317 HFEAEDEIPTIVESSHDDQAAS------PVPSEEDVHEQIQTTTTTTTVTKEHFVPDDEI 2370

Qy   1462 QRERSEEEDKPDAGPDADGK-------ESTAIEKVLKA--EPE-------------ILEL 1499

Db   2371 DSEHMNESDKYASGSPVPSEEDSSRVTETTTTTTVTREHFEPEDDHSPVVQTQEYSASES 2430

Qy   1500 ESKSNKIVLNVIQTAVDQFARTETAPETHAYD----------SQTQVPACRLDSREPNRC 1549

Db   2431 PVPSEKSVERVIETTT-----TTTVTREHFEDEDHILGQQGESDSQIPSESITSENMDR- 2484

Qy   1550 WTKMKDAKMKHPVPQPREDLQVLTVLEAWAQP 1581

Db   2485 ------ETSSSPVQSNRDEEFVLPAIAPYKQP 2510
```

RESULT 10
T26216
hypothetical protein W06A7.3c - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26216
R;Ainscough, R.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z20173
A;Accession: T26216
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2484 <WIL>
A;Cross-references: EMBL:Z78066; PIDN:CAB51467.1; GSPDB:GN00023; CESP:W06A7.3c
A;Experimental source: clone W06A7
C;Genetics:
A;Gene: CESP:W06A7.3c
A;Map position: 5
A;Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2

```
  Query Match             6.0%;  Score 485;  DB 2;  Length 2484;
  Best Local Similarity   20.2%;  Pred. No. 5.8e-10;
  Matches  362;  Conservative  244;  Mismatches  648;  Indels  538;  Gaps   69;

Qy    191 AEASVG---AGDH-QEPSVETAVGESAS--KESELKQSTEKQEGTLKQEQSSTEIPLQAE 244

Db      3 APAVIGLEDSGDNLNRPALETASDQESSENKEPDASENVEETEPV----SSAKVIALESS 58

Qy    245 SD--------QAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWR 296

Db     59 GDCEEGNIKISANENSVEPDGADKPAH---TEQEIPTIILPSETTVT------------ 102
```

```
Qy   297 KKTSFKKSKEDDL-ETAEKRKEQEAEKVDE------EEKEKTEPASEEQEPAEDTD-QAR 348
Db   103 -------QLEDHLSETEEKQNEVESSIKSEKNVIGLEDSGDTFRDTSGLAPAEDNEAEAT 155

Qy   349 LSADYEKVELPLEDQVGDLEASSEEKCAPLATEV--FDEKMEAH--------------- 390
Db   156 ITTDF----VPLE-SAGDIPSENEIKEVASAPDVVGLEEYIIGNIPNAPVVNDDIPNVFT 210

Qy   391 QEVVAEVHVSTVEKTEEEQGGGGEAEGGVVV--EGTGESL----PPEKLAEPQE-VPQEA 443
Db   211 PEVANDETVETFSVTAE------EASIPVVVELEPIGDEYEFQRPVENFSEPSDNINLEE 264

Qy   444 EPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEM-----------LSS 491
Db   265 SGAEQVLLENNM------FTPLDELEPQQKILNEKAEHI--EIEASGDEFIKDHPFPVEN 316

Qy   492 QERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPE------SA 545
Db   317 EQNVNEQSSRVETVHSFIGLES-SG---VGLTGAVSDSVANNVKE-NTESPDIISLEASG 371

Qy   546 D-------------EQKGESSASSPEEPEETTCL---------------------EKGP 570
Db   372 DELSKLVEAREIITESKDAYSTDVPESRKTVSDVIGLEEAGDKIVSNNVSNVMGNPDESQ 431

Qy   571 LEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKE----EELEKVK 626
Db   432 LALEQADNVPELPIENSEQETVAVKATESCDHVVDSQKTLERASSLEEDIMSPEVLGITS 491

Qy   627 SATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSD 686
Db   492 SQTLSDYLPVISEDQDSIPV---------PVTEVEE-----------TSEKLVKASSLEE 531

Qy   687 DEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTW 746
Db   532 DVVSPEVLELDD---RVQNKNPESEVTAVDASKTEGD---FSDSPDSRAT---------- 575

Qy   747 ESF-KRLVTPR--------KKSKSKLEEKAEDSSVEQ-------LSTEIEPSRE------ 784
Db   576 ETFMEKLVTVTENLLPAGDKLSEERIQEIRENETISQPGKEEDDLENANDPDDETIVEKI 635

Qy   785 ----ESWVSIKKFI-------PGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLS 833
Db   636 VSMAESSLPIEAVISTEDGGTSDQPAQNAIPDSEETTVDDSQTEEIFTDDN-----VKKS 690

Qy   834 EYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSW 893
Db   691 KENT-----PKAENDTEINYLPGGEEGPED-----------------NAEKRNEAVSPND 728

Qy   894 ISASVTEPLEHTAGEAMPP----------VEEVTEKDIIAEETPVLTQTLPEGKDAHDDM 943
Db   729 ETSEIKQDLENLENGASGPDNNVQVDEAAQEDPTDPETVDETTSKISENMPKAPDTEDDN 788

Qy   944 VTSEVDFTSEAVTATETSEAL----------------------------------RTEEV 969
Db   789 ATEFIE------AGLETTEAFGDAEHVSYLDANIEKLVAMADEPLPVDELVSIEERPEEV 842

Qy   970 TEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLD-----TEEERQTQAILQA 1024
Db   843 APAESTGEDEDIPRRDRRTVSLTGTGDQNAPIQVIFVGDGDENPDANADQERTSEHNELI 902

Qy  1025 VADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSE---------------------VL 1062
Db   903 ESDKESEEAITKNEEDVDQDPIQSEEPLTSQEGESSIGNKIVAVVGSVLLGGAVIPYGVL 962

Qy  1063 ASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHV--ATCQVIKLQQ 1120
Db   963 ASNENEDAHADREVEET-------GDSTRDRPEEETFVSKLTSMVENILPSTNDENPEAV 1015

Qy  1121 LMEQAVAPESSETLTDS-ETNGSTPLAD-------------SDTADGTQQDETI------ 1160
Db  1016 SMVENVLPVNTEGLDESKEDNPDAPTAEAHSGEKNLRNDKTTDTRERDPEEETILNKLVE 1075

Qy  1161 -----------------DSQDSKATAAVRQSQVTEEEAATAQKEE------------PS 1190
Db  1076 NALPTGVTGSFTEVSAPDAQELDETVVDHAGQNDTSEVEDAPEKSAGGTVIEKFTSMIES 1135

Qy  1191 TLPNNVPAQEEHGEE------PGRDVLEPTQQELTAAAVPVLAKTEVG----QEGEVDWL 1240
Db  1136 ILPVQAPTQPENAESHIDQETGGASEIKDDQNQPEEFSAEHQGKFEVSAEPDQESAEFQL 1195

Qy  1241 DGEKVKEEQEVFVHSGPNSQKAAD-------VTYDSEVMGVAGCQEKESTEVQSLSLEEG 1293
Db  1196 EAKKDQDKETI-----ENSEDAKKETVMEKLVSLVENILPVEAVLPSDST-VTKNSEDKK 1249

Qy  1294 EMETDVEKEKRETK----PEQVSEEGEQETAAPE-----HEGTYGKPVLTLDMPSSERGK 1344
Db  1250 ELETQ-ELSSKEIKTSGQPEYVPETSEAFVSDPEIFQRVKRASSTEP----KTQKTEPHA 1304

Qy  1345 ALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH 1404
Db  1305 PIFIVGQSTEDDEQSIANVIDELVHEDD------EKKVPEVTANISVSASENIDDSTTAN 1358

Qy  1405 LLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQR- 1463
Db  1359 AVP--KTEVSSEQLQVATVEFELESAPEEESAAIPEVQEPLEKVEVQPDLSQNSPAPHKI 1416

Qy  1464 -----------------------ERSEEEDKPDAGPDADGKESTAIE------------- 1487
Db  1417 IDLHFNIPKDHEDYGNDYVPFGTESSEESQKADGNQENQEEEDVVAELNFHPIRQWRDED 1476

Qy  1488 -------------------------------------KVLKAEPEILELESKSNKIVLNV 1510
Db  1477 VISLQSLKSLVAEVGCITDVDASDVNEQDEESTLKILKVVPSEPSLLELDFTNDPKVIHV 1536

Qy  1511 ------------------IQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWT 1551
Db  1537 PIPLMEPATMYLEEMVEWIIADAVKEVSEMEVVTESEISEMAPQVSESTCPIPEP----- 1591

Qy  1552 KMKDAKM------KHPVPQPREDLQVL-TVLEAWAQPRKCLPRLQLKAPVSK 1596
Db  1592 -LADLKLPVEDDEKTPEPEPVVPGQVQERIIPIEVEQAPTIPQRPPRAPKSE 1642
```

RESULT 11
T26215
hypothetical protein W06A7.3a - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T26215
R;Ainscough, R.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z20173
A;Accession: T26215
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-2607 <WIL>
A;Cross-references: EMBL:Z78066; PIDN:CAB01522.2; GSPDB:GN00023; CESP:W06A7.3a
A;Experimental source: clone W06A7
C;Genetics:
A;Gene: CESP:W06A7.3a
A;Map position: 5
A;Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2; 2586/2

```
  Query Match             6.0%;  Score 485;  DB 2;  Length 2607;
  Best Local Similarity   20.2%;  Pred. No. 6.1e-10;
  Matches  362;  Conservative  244;  Mismatches  648;  Indels  538;  Gaps   69;

Qy   191 AEASVG---AGDH-QEPSVETAVGESAS--KESELKQSTEKQEGTLKQEQSSTEIPLQAE 244
Db     3 APAVIGLEDSGDNLNRPALETASDQESSENKEPDASENVEETEPV----SSAKVIALESS 58

Qy   245 SD--------QAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWR 296
Db    59 GDCEEGNIKISANENSVEPDGADKPAH---TEQEIPTIILPSETTVT------------ 102
```

```
Qy    297 KKTSFKKSKEDDL-ETAEKRKEQEAEKVDE------EEKEKTEPASEEQEPAEDTD-QAR 348
Db    103 -------QLEDHLSETEEKQNEVESSIKSEKNVIGLEDSGDTFRDTSGLAPAEDNEAEAT 155

Qy    349 LSADYEKVELPLEDQVGDLEASSEEKCAPLATEV--FDEKMEAH---------------- 390
Db    156 ITTDF----VPLE-SAGDIPSENEIKEVASAPDVVGLEEYIIGNIPNAPVVNDDIPNVFT 210

Qy    391 QEVVAEVHVSTVEKTEEEQGGGGEAEGGVVV--EGTGESL----PPEKLAEPQE-VPQEA 443
Db    211 PEVANDETVETFSVTAE------EASIPVVVELEPIGDEYEFQRPVENFSEPSDNINLEE 264

Qy    444 EPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEM-----------LSS 491
Db    265 SGAEQVLLENNM------FTPLDELEPQQKILNEKAEHI--EIEASGDEFIKDHPFPVEN 316

Qy    492 QERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPE------SA 545
Db    317 EQNVNEQSSRVETVHSFIGLES-SG---VGLTGAVSDSVANNVKE-NTESPDIISLEASG 371

Qy    546 D-------------EQKGESSASSPEEPEETTCL---------------------EKGP 570
Db    372 DELSKLVEAREIITESKDAYSTDVPESRKTVSDVIGLEEAGDKIVSNNVSNVMGNPDESQ 431

Qy    571 LEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKE----EELEKVK 626
Db    432 LALEQADNVPELPIENSEQETVAVKATESCDHVVDSQKTLERASSLEEDIMSPEVLGITS 491

Qy    627 SATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSD 686
Db    492 SQTLSDYLPVISEDQDSIPV---------PVTEVEE-----------TSEKLVKASSLEE 531

Qy    687 DEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTW 746
Db    532 DVVSPEVLELDD---RVQNKNPESEVTAVDASKTEGD---FSDSPDSRAT---------- 575

Qy    747 ESF-KRLVTPR---------KKSKSKLEEKAEDSSVEQ-------LSTEIEPSRE------ 784
Db    576 ETFMEKLVTVTENLLPAGDKLSEERIQEIRENETISQPGKEEDDLENANDPDDETIVEKI 635

Qy    785 ----ESWVSIKKFI-------PGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLS 833
Db    636 VSMAESSLPIEAVISTEDGGTSDQPAQNAIPDSEETTVDDSQTEEIFTDDN-----VKKS 690

Qy    834 EYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSW 893
Db    691 KENT-----PKAENDTEINYLPGGEEGPED----------------NAEKRNEAVSPND 728

Qy    894 ISASVTEPLEHTAGEAMPP---------VEEVTEKDIIAEETPVLTQTLPEGKDAHDDM 943
Db    729 ETSEIKQDLENLENGASGPDNNVQVDEAAQEDPTDPETVDETTSKISENMPKAPDTEDDN 788

Qy    944 VTSEVDFTSEAVTATETSEAL----------------------------------RTEEV 969
Db    789 ATEFIE------AGLETTEAFGDAEHVSYLDANIEKLVAMADEPLPVDELVSIEERPEEV 842

Qy    970 TEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLD-----TEEERQTQAILQA 1024
Db    843 APAESTGEDEDIFRRDRRTVSLTGTGDQNAPIQVIFVGDGDENPDANADQERTSEHNELI 902

Qy   1025 VADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSE---------------------VL 1062
Db    903 ESDKESEEAITKNEEDVDQDPIQSEEPLTSQEGESSIGNKIVAVVGSVLLGGAVIPYGVL 962

Qy   1063 ASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHV--ATCQVIKLQQ 1120
Db    963 ASNENEDAHADREVEET-------GDSTRDRPEEETFVSKLTSMVENILPSTNDENPEAV 1015

Qy   1121 LMEQAVAPESSETLTDS-ETNGSTPLAD------------SDTADGTQQDETI------ 1160
Db   1016 SMVENVLPVNTEGLDESKEDNPDAPTAEAHSGEKNLRNDKTTDTRERDPEEETILNKLVE 1075

Qy   1161 -----------------DSQDSKATAAVRQSQVTEEEAATAQKEE-------------PS 1190
Db   1076 NALPTGVTGSFTEVSAPDAQELDETVVDHAGQNDTSEVEDAPEKSAGGTVIEKFTSMIES 1135

Qy   1191 TLPNNVPAQEEHGEE------PGRDVLEPTQQELTAAAVPVLAKTEVG----QEGEVDWL 1240
Db   1136 ILPVQAPTQPENAESHIDQETGGASEIKDDQNQPEEFSAEHQGKFEVSAEPDQESAEFQL 1195

Qy   1241 DGEKVKEEQEVFVHSGPNSQKAAD-------VTYDSEVMGVAGCQEKESTEVQSLSLEEG 1293
Db   1196 EAKKDQDKETI-----ENSEDAKKETVMEKLVSLVENILPVEAVLPSDST-VTKNSEDKK 1249

Qy   1294 EMETDVEKEKRETK----PEQVSEEGEQETAAPE-----HEGTYGKPVLTLDMPSSERGK 1344
Db   1250 ELETQ-ELSSKEIKTSGQPEYVPETSEAFVSDPEIFQRVKRASSTEP----KTQKTEPHA 1304

Qy   1345 ALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAH 1404
Db   1305 PIFIVGQSTEDDEQSIANVIDELVHEDD------EKKVPEVTANISVSASENIDDSTTAN 1358

Qy   1405 LLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQR- 1463
Db   1359 AVP--KTEVSSEQLQVATVEFELESAPEEESAAIPEVQEPLEKVEVQPDLSQNSPAPHKI 1416

Qy   1464 ----------------------ERSEEEDKPDAGPDADGKESTAIE------------- 1487
Db   1417 IDLHFNIPKDHEDYGNDYVPFGTESSEESQKADGNQENQEEDVVAELNFHPIRQWRDED 1476

Qy   1488 ------------------------------------KVLKAEPEILELESKSNKIVLNV 1510
Db   1477 VISLQSLKSLVAEVGCITDVDASDVNEQDEESTLKILKVVPSEPSLLELDFTNDPKVIHV 1536

Qy   1511 ------------------IQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWT 1551
Db   1537 PIPLMEPATMYLEEMVEWIIADAVKEVSEMEVVTESEISEMAPQVSESTCPIPEP----- 1591

Qy   1552 KMKDAKM------KHPVPQPREDLQVL-TVLEAWAQPRKCLPRLQLKAPVSK 1596
Db   1592 -LADLKLPVEDDEKTPEPEPVVPGQVQERIIPIEVEQAPTIPQRPPRAPKSE 1642
```

RESULT 12
S00485
gene 11-1 protein precursor - malaria parasite (Plasmodium falciparum) (fragments)
C;Species: Plasmodium falciparum
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 09-Jun-2000
C;Accession: S00485
R;Scherf, A.; Hilbich, C.; Sieg, K.; Mattei, D.; Mercereau-Puijalon, O.; Mueller-Hill, B
EMBO J. 7, 1129-1137, 1988
A;Title: The 11-1 gene of Plasmodium falciparum codes for distinct fast evolving repeats
A;Reference number: S00485; MUID:88296416; PMID:2841111
A;Accession: S00485
A;Molecule type: DNA
A;Residues: 1-1315;1316-1485;1486-1657;1658-1729;1730-1948 <SCH>
A;Cross-references: EMBL:X07453
C;Comment: This protein is associated with the membrane of red blood cells at the schizo
C;Genetics:
A;Gene: 11-1
A;Introns: 71/3
C;Keywords: tandem repeat
F;1-71/Domain: signal sequence #status predicted <SIG>
F;72-1948/Product: gene 11-1 protein (fragments) #status predicted <MAT>

```
  Query Match            6.0%;  Score 483.5;  DB 2;  Length 1948;
  Best Local Similarity  21.9%;  Pred. No. 4.9e-10;
  Matches  371;  Conservative  252;  Mismatches  619;  Indels  455;  Gaps  76;

Qy     28 GHGPAAEASGAAGDPADADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDE 87
Db    283 GENEGGDGHGSAGDSAGGD-------RGGDENKDKDNEENKDKKGNKDDDDEEEEEEEED 335

Qy     88 DVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEI---------IEQIPASENNVE 139
Db    336 DEEKEEKDDKDEKDEKDEKDEKDEKDEKDEKDEKGEKDESKGEEKFLKEEKMPKKEENMQ 395
```

```
Qy    140 EMVQPAESQANDVGFKKVFKFVGFK-------------FTVKKDKNEKSDTVQLLTVKKD 186
Db    396 DKIERDK------------EFEGFKDEYFTQMEGGAYGITVEHEGQKLRETLT------- 436

Qy    187 EGEGAEASVGAGDHQEPS------VETAVGESASKESEL-----------------KQS 222
Db    437 --EGVEGYEKVIEEMEQTKLDEFLLPTNEGKSDDFVSEMEEERLFNFVSYMEEEVKDKMF 494

Qy    223 TEKQEGTLKQ-------------EQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSP 269
Db    495 IEREQGTEEEVIHEDLSDEKIGIEKKELETFVQDVSDEVLQDEQLIEEEKLTEEEKLTEE 554

Qy    270 ESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEK- 328
Db    555 EKLTEE--------------------EKLTEEEKLTEDE-QLIEEEKLTEEEKLTEEEKL 593

Qy    329 EKTEPASEEQEPAED---TDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDE 385
Db    594 TEEEKLTEEEKLTEEEKLTEEEKLTEEEIFIE---EDKI-----SPKEVEEPL-EDVLME 644

Qy    386 KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAE---------- 435
Db    645 KMEKEYKEILHEGKSVVTQDMDEKLLSENLEDDLIPEEIIEELVDEELVEEVTAEEIVEE 704

Qy    436 -----------PQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVS 484
Db    705 FVVEVEIEEKIPKELVEEVMPEED----------------VDEVVPEE--LVEEPDEVVE 746

Qy    485 EVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGE-YQHIHTESPE 543
Db    747 EVTTEELTEEV-IPEELIEKVVEEELVEKVVAQKIV-------EVIPEELIEEVITEKP- 797

Qy    544 SADEQKGE-SSASSPEE------PEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITP 596
Db    798 -AEEVIPELVEEVIPEEIVEEVLPEEVV-KEVIPEEVVEEVIPEEVVEEVPEEIVEEVVP 855

Qy    597 WASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEP 656
Db    856 ----EELV---------EEMKPEEVFEEVISEEL------VEEMKPE--EVVEEVTPE-- 892

Qy    657 KRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVP 716
Db    893 --KVVEEVVPEELV--------------------------------EEMKPEEVVEEVVP 918

Qy    717 ASTQEQ---DQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKK-----SKSKLEEKAE 768
Db    919 EELVEEVIPEEVVEEIIPEEVVHEAKHE-----ELFEKLV-PEVEVEEIIPKELVEVIPE 972

Qy    769 DSSVEQLSTEIEPSR--EESWVS---IKKFIPGRRKKRA---DGKQE---QATVEDSGPV 817
Db    973 EVLPEELIEEVIPEEIIEEVVIPEEVVEEVIPEEIVQEVIPEDLMEEVVPEEIIEEKIPE 1032

Qy    818 EINEDDPN---VPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLV----- 869
Db   1033 EIVEEEIHEEIVEEIIP----EEVVEEVVPAEVEEMIPEKIVEEVIPEELVEEVVPVEVL 1088

Qy    870 -HTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPP--VEEVTEKDIIAEET 926
Db   1089 EEVVPVEVLEEVIPEEVVEEVIPVEVLEEVIP--EEVVEEVIPEEFIEEMIPEEIVEEVV 1146

Qy    927 P-VLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAV 985
Db   1147 PEVVEEVIPE--ELVEEVVPVEL---LEEVISEELLEEVIPEELVEEVVPEELVEEMKPE 1201

Qy    986 SQLTD-SPDT-TEEATP---VQEVESGVLDTEE----------------------EERQ 1017
Db   1202 EVLEEVIPEVLVEEVIPEELVEEVIPEVLVEEEVPEELVEEVIPELLVEEVVPEVVEEVQ 1261

Qy   1018 TQAILQAVADKVKEE-------SQVPATQTVQRTGSKALEKV--EEVEED---------S 1059
Db   1262 EQLVEEFIPEEVVEELVPEELVEEVIPEEIVEEVVEEVVEEIIPEEVIEEVRPGPEELVE 1321

Qy   1060 EVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESL---EVPEVTADVDHVATCQVI 1116
Db   1322 EVIPEELVEEVIPEELVAEVIPEELVE-----EVIPEELVEEVVPEVVEEVVPEELVEEV 1376

Qy   1117 KLQQLMEQAV--------APES--SETLTDSETNGSTP------LADSDTADGTQQDETI 1160
Db   1377 IPEELVEEVIPEVLVEEEVPEELVEEVIPEELVEEMLPEELIEVVVPEELVEEVVPEELV 1436

Qy   1161 DSQDSKATAAVRQSQVTEE-------EAATAQKEEPSTLPNNVPAQEEHGEE--PGRDVL 1211
Db   1437 EE--------VVPEEVVEEVIPEELIEELVPEVEIEEILP------EELIEELIPEEIIE 1482

Qy   1212 EPTQQELTAAVPVLAKTEVGQEGEVDWLDGEKVKEE-------QEVFVHSGPNSQKAAD 1264
Db   1483 EVIPEELVAEVIPGEIVEEVVPEELVEEVVPEEVVEEVIPKEIVEEVV-----SEEVVE 1536

Qy   1265 VTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEH 1324
Db   1537 EVIPKEI-----AEEVLSAQVFEEVIPKEIVEEEVPEELVEVAPEELVEE-----VVPEE 1586

Qy   1325 EGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEK 1384
Db   1587 --------LSEEVVPEELVEEV--------VPE----GLVE-EVVPEELVVFLPEELVEE 1625

Qy   1385 VIETVVISETGES------PECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESI 1438
Db   1626 VVPEVV-EEVEEELVEEFIPEEVVEELVPEE------------------LVPEKLIEDV 1665

Qy   1439 PIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILE 1498
Db   1666 ------VPEEIVEEVLPEEV---VKEVIPEEVVEEVLPEELVEEVVPEELVEEVIPEELI 1716

Qy   1499 LESKSNKIVLNVIQTAV 1515
Db   1717 AEVVSEQLVEEVLPEEV 1733
```

RESULT 13
A56577
microtubule-associated protein MAP 1B - rat (fragment)
C;Species: Rattus norvegicus (Norway rat)
C;Date: 21-Jul-1995 #sequence_revision 21-Jul-1995 #text_change 16-Feb-1997
C;Accession: A56577
R;Zauner, W.; Kratz, J.; Staunton, J.; Feick, P.; Wiche, G.
Eur. J. Cell Biol. 57, 66-74, 1992
A;Title: Identification of two distinct microtubule binding domains on recombinant rat M
A;Reference number: A56577; MUID:92347374; PMID:1639092
A;Accession: A56577
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-2364 <ZAU>
A;Cross-references: GB:X60550
A;Experimental source: brain
A;Note: nucleotide sequence not given; conceptual translation not complete
C;Superfamily: microtubule-associated protein MAP1B

```
  Query Match            5.8%;  Score 470;  DB 2;  Length 2364;
  Best Local Similarity   20.2%;  Pred. No. 1.9e-09;
  Matches  328;  Conservative  234;  Mismatches  602;  Indels  458;  Gaps   62;

Qy     45 ADPATKL-----PQKNGQLSSVNGVAEQGDVH------VQEENQEGQEEEVVDEDVGQRE 93
Db    382 ANPAEKIIRVLFPGNSTQYNILEGLEKLKHLDFLKQPLATQKDLTGQVSTPPVKQVKLKQ 441

Qy     94 SEDVRE--KDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQAND 151
Db    442 RADSRESLKPATKPLSSKS-----VRKESKEEAPE------ATKASQVEKTPKVESKE-- 488

Qy    152 VGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGE 211
Db    489 ------------KVIVKKDKPGKVESKPSVT---------EKEVPSKEEQSP-VKAEVAE 526

Qy    212 SASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPES 271
```

```
Db    527 KAATESKPKVTKDK---VVKKE-------IKTKPEEKKEEKPKKEVAKKEDKTPLKKDEK 576

Qy    272 PSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEK--E 329

Db    577 PK---KEEAKKEIKK------------EIKKEEKKELK-KEVKKETPLKDAKKEVKKDE 619

Qy    330 KTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEA 389

Db    620 KKEVKKEEKEPKKEI--KKISKDIKK-STPLSD---------TKKPAALKPKVAKKEEPT 667

Qy    390 HQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEEL 449

Db    668 KKEPIAAGKLKDKGKVKVIKKEGKTTEAAATAVGTAAVAAAAGVA-------ASGPAKEL 720

Qy    450 MKSREMCVSGGDHTQ-LTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSS 508

Db    721 EAERSLMSSPEDLTKDFEELKAEEIDVAK---DIKPQLELIEDEEKLK------------ 765

Qy    509 SGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEK 568

Db    766 ------------------ETEPGEAYVIQKE----------------------TEVSK 783

Qy    569 GPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSA 628

Db    784 GSAESPDEGI----TTTEGEGECE-----------QTP-------------EELEPV--- 812

Qy    629 TLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDE 688

Db    813 --------------------EKQGVDDIEKFEDEGAGFEESSEAGDYEEKA--------- 843

Qy    689 GGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWES 748

Db    844 -----------ETEEAEEPEEDG----------EDNVSGSASKH---SPTEDEEIAKAEA 879

Qy    749 FKRLVTPRKKSKSKLEEKAEDSSVEQLST-EIEPSREESWVSIKKFIPGRRKKRADGKQE 807

Db    880 DVH-IKEKRESVASGDDRAEEDMDEALEKGEAEQSEEEG-------------EEEEDKAE 925

Qy    808 QATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKT 867

Db    926 DAREEDHEPDKTEAED-YVMAVVDKAAEAGVTEDQYDFLGT---PAKQPGVQSPSREPAS 981

Qy    868 LVHTVSVAVIDGTRAVTSVEER---SPSWISAS--VTEPLEHTAGEAMPPVEEVTEKDII 922

Db    982 SIHDETLPGGSESEATASDEENREDQPEEFTATSGYTQSTIEISSEPTPMDEMSTPRDVM 1041

Qy    923 AEET----------------------------PVLT--QTLPEGK--DAHDDMVTSEVD 949

Db   1042 TDETNNEETESPSQEFVNITKYESSLYSQEYSKPVVASFNGLSDGSKTDATDGR-----D 1096

Qy    950 FTSEAVTATETS-------------EALRTEEVTEASGAEETTDMVS-------AVSQLT 989

Db   1097 YNASASTISPPSSMEEDKFSKSALRDAYRPEETDVKTGAELDIKDVSDERLSPAKSPSLS 1156

Qy    990 DSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKAL 1049

Db   1157 PSPPSPIEKTPLGERSVNFSLTPNE--------------IKASAEGEATAVVSPGVTQAV 1202

Qy   1050 ---------EKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEV 1100

Db   1203 VEEHCASPEEKTLEVVSPSQSVTGSAGHTPYYQSPTDEKSSHLPTEVTENAQAVPVSFEF 1262

Qy   1101 PEVTADVDHVATC----QVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQ 1156

Db   1263 TEAKDENERSSISPMDEPVPDSESPIEKVLSPLRSPPLIGSESAYEDFLSADDKALGRRS 1322

Qy   1157 DETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQ 1216

Db   1323 ESPFEGKNGKQGFSDKESPVSDLTSDLYQDKQEEKRAGFIPIKEDFSPEKKASDAEIMSS 1382

Qy   1217 ELTAAAVPVLAKTEVGQEG---EVDWLDGEKVKEEQEVFVHSGPNSQKA-------ADVT 1266

Db   1383 QSALA----LDERKLGGDGSPTQVDVSQFGSFKEDTKMSISEGTVSDKSATPVDEGAEDT 1438

Qy   1267 YDSEVMGVAGCQ-------------------------EKESTEV-QSLSLEEGE----- 1294

Db   1439 Y-SHMEGVASVSTASVATSSFPEPTTDDVSPSLHAEVGSPHSTEVDDSLSVSVVQTPTTF 1497

Qy   1295 METDVEKEKRE-------TKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALG 1347

Db   1498 QETEMSPSKEECPRPMSISPPDFSPKTAKSRTPVQDHRSEQSSMSIEFGQESPEHSLAMD 1557

Qy   1348 SLGGSPSLPDQDKAGCI---------------EVQVQSLDTTVTQTAEAV----EKVIET 1388

Db   1558 FSRQSPDHPTVG-AGMLHITENGPTEVDYSPSDIQDSSLSHKIPPTEEPSYTQDNDLSEL 1616

Qy   1389 VVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPES 1448

Db   1617 ISVSQVEASPSTSSAH-TPSQIAS---------------PLQEDTLSD-----VVPPRDM 1655

Qy   1449 TLHPDLQGE-ISASQRER----------SEEEDKPDAGP-----DADGKESTAIEKVLKA 1492

Db   1656 SLYASLASEKVQSLEGEKLSPKSDISPLTPRESSPTYSPGFSDSTSGAKESTAAYQTSSS 1715

Qy   1493 EP 1494

Db   1716 PP 1717
```

RESULT 14
D96796
probable heat shock protein, 53413-59028 [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: D96796
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: D96796
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1871 <STO>
A;Cross-references: GB:AE005173; NID:g6143906; PIDN:AAF04452.1; GSPDB:GN00141
C;Genetics:
A;Gene: F28O16.15
A;Map position: 1

```
  Query Match             5.8%;  Score 466.5;  DB 2;  Length 1871;
  Best Local Similarity   20.5%;  Pred. No. 1.9e-09;
  Matches  331;  Conservative  248;  Mismatches  588;  Indels  447;  Gaps   68;

Qy     50 KLPQKNGQLSSVNGVAEQGDVHVQEENQE---------------------GQEEEVVDED 88

Db    310 KLQEKEEQHSK----EQKG--HSKEENMKELVEEKTPEAETTIRNDILGPGQEIEVPEVD 363

Qy     89 VGQRESEDVREKDR-VEEMAANSTAVEDI-TKDGQEETSEIIEQIPASE---------N 136

Db    364 TLGKTSDEGKEKQNIVKKEIKNGDATEEIDAKMGEVFASNIADTGMNSEDFESDKLESAD 423

Qy    137 NVEEMVQPAE---------SQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDE 187

Db    424 EVDKMVEKKDRQEENDKVGAQSEDISLTKLQEIGEQQFQGQK-RHDKQENIKELR----E 478

Qy    188 GEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEK--QEGTLKQEQSSTEIPLQAES 245

Db    479 GQASEA--------EKNIKNDILKPVQKRSEGKHKIQKTFQEETNKQPEGYNEKIM--ET 528
```

```
Qy    246 DQAAEEEAKDEGEE---KQE-KEPTKSPESPSSP--VNSETTSSFKKFFTHGWAGWRKKT 299
Db    529 GKKINEDGTRKVQEMIRQQELDEPARSEKENRSRELVKSKTNDEEKK---------EKEI 579

Qy    300 SFKKSKEDDLETAEKRKEQE-AEKVDE---------EEKEKTEPASEEQEPAEDTDQARL 349
Db    580 AGTERKEKESDRPKILREQEVADEVAEDKTKFSIYGEVKEEEEIAGKEKEFGSDDDIARI 639

Qy    350 SADYEKV---------------ELPLEDQVGD--------LEASSEEKCAPLATEVFDEK 386
Db    640 VRDTEQLDSNAMKGQEEKDMIQELVLEEKVCDGGKGIIAVAETKAENNKSKRVQETEEQK 699

Qy    387 ME------AHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVP 440
Db    700 LDKEDTCGKHFQKLIEGEISDHGEVEDVEKGKKRTEA-------------------EKRIK 741

Qy    441 QEAEPAEELMKSREMCVSGG--DHTQLTDL-----------SPEEKTLPKHPEGIV---- 483
Db    742 DRAREAEEI-KEKDLGVSGRYIKGTTIKELVENRGIYRNEHEEKKKDDANRPEKITGTIK 800

Qy    484 SEVEMLSSQ-----------------ERIK----VQGSPLKKLFSSSGLKKLSGKKQ--- 519
Db    801 QELVSLNSQLRQENVEDGDKTQELVEEKIKDCEEEEGSEESKIKTDDVVRKVQGIKEEEL 860

Qy    520 -KGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGE 578
Db    861 YKPKREHG-----------TKITELVEETTGD---YEKQEEKET---------AESDIE 896

Qy    579 AEEGTTS--DGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEK-VKSATLSSTDS 635
Db    897 AECGSLRKVDGIEEHELHEPKIHKERDNNRVTGAKEPSGQEKGEKEEKIVESMTITENDN 956

Qy    636 TVSEMQDEVKTVG-------------------EEQKPEEPKRRVDTSVSWEALICVGSS 675
Db    957 SIDVQETKKERPGRLESHDKRYKIQELLMEAGHNDRKEEEQNENVTAEVELET---ERVS 1013

Qy    676 KKRARKASSSDDEGG---------------------------------PRTLGGDSHRA 701
Db   1014 SKKVQEGKMEDDNSGKFHEFEERKSYEDWTHEKREKRKVLVEEEETYPKDKHTGGEDHND 1073

Qy    702 EEASKDKEAGTDAVPASTQE-----------QDQAQGSSSPEPAGSPSEGEGVSTWESFK 750
Db   1074 HKEEEQKENVIAKAELNTEEDSFKKVEEIEKQDHGELKRSMVQAKRQETEEKDKTRAMEK 1133

Qy    751 RLVTPRKK-----SKSKL-----------EEKAEDSSVEQL-STEIEPSRE------ESW 787
Db   1134 NETVERRKQTKDGSLGKLREGEDPELGGHERRGEEDRIEELVETEISDHKEKVKKKDEDY 1193

Qy    788 V----SIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSE------YNA 837
Db   1194 ILRSQDTGKVDLGERERRSKQRKIHKSVEDEIGDQEDEDAEEAAAVVSRNENGSSRKVQT 1253

Qy    838 VEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISAS 897
Db   1254 IEEESEKHKEQNKIPETSNP-EVNEE--------------DEERVV----EKETKEVEAH 1294

Qy    898 V------TEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFT 951
Db   1295 VQELEGKTENCKDDDGEGRR--EERGKQGMTAEN--MLRQRF---KTKSDDGIVRKIQET 1347

Qy    952 SEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDT 1011
Db   1348 KE-------------EEPDEKKSQESSSHVVKLVAE--DG-----------SLRNG-LEF 1380

Qy   1012 EEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEK-------VEEVEEDSEVLAS 1064
Db   1381 SEKESTVSKMLKLDESKEKEEHKKIRKPTEERSNAPVIEKQGNKKNAEEEMQDKIDRRGK 1440

Qy   1065 EKE-KDVMPKGPVQEAGAEHLAQ---GSETGQATPESLEVPEVTADVDHVATCQVIKLQQ 1120
Db   1441 NQEIKGQEPYGVLRNGEHDKITEYHRGEEKGTA---------------ENVSS---TKIQQ 1483

Qy   1121 LMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEE 1180
Db   1484 TKDELEKPRKPSEISENHN-------IHEFMDSSQSQDIEEKGSDQAEKYAKQNKIQE-- 1534

Qy   1181 AATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQ------E 1234
Db   1535 -----------VMNDEDKKEEY------HISERVRNEMAKRILQVESKANDGSSKKNETE 1577

Qy   1235 G-EVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEG 1293
Db   1578 GQESTGLRGRKKRENHQELVELETSDQKKG--VKEDEVVGKAEIIEDEYDSSRKIHEHEE 1635

Qy   1294 EMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSL---- 1349
Db   1636 RMSDKLEMHGEEEMSEKLAEEETSDGEEAKEGNRAGK---------KSRDDGFGKVRKIE 1686

Qy   1350 ----GGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPE 1399
Db   1687 VQRKDNDQSFVEKDTSGKAKENLNDEEPTKTETKATDNESRKIHQIKEQGTSEQ 1740
```

RESULT 15
T34518
nestin - golden hamster (fragment)
C;Species: Mesocricetus auratus (golden hamster)
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C;Accession: T34518
R;Steinert, P.M.; Chou, Y.H.; Prahlad, V.; Parry, D.A.D.; Marekov, L.; Wu, K.C.; Jang, S
submitted to the EMBL Data Library, December 1998
A;Description: A high molecular weight intermediate filament associated protein in BHK-2
with type III vimentin and type IV a-internexin.
A;Reference number: Z21538
A;Accession: T34518
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1804 <STE>
A;Cross-references: EMBL:AF110498; NID:g4063502; PID:g4063503; PIDN:AAC98312.1
A;Experimental source: cell line BHK-21

```
  Query Match             5.7%;  Score 456.5;  DB 2;  Length 1804;
  Best Local Similarity   20.8%;  Pred. No. 4.2e-09;
  Matches  358;  Conservative  249;  Mismatches  575;  Indels  541;  Gaps   79;

Qy      2 GAGSSTEQRSPEQP------AGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKN 55
Db    246 GVLTELEEAGGQQPGHFPEDATASAPS---LSPHHPVLEAKD--GDSTES---------- 290

Qy     56 GQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVED 115
Db    291 -----------RGSSIFQED--EGQIWELV-------------EKEAAIELKVESSLAQE 324

Qy    116 ITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKS 175
Db    325 TQEDGLH-----TEEIQDSQGPLQK----------------------------------- 344

Qy    176 DTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVG-ESASKESELKQSTEKQEGTLKQEQ 234
Db    345 ETLEAL------GEEPLMSLKIQNHETPGKENCNSLRSVDENQGTLKSPEEEKQTLLKSL 398

Qy    235 SSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAG 294
Db    399 EEKDVEVEKTLEKGVPELSKPLGKE----DPRIEDQELMSPEGTLETLSF--------IG 446

Qy    295 WRKKTSFKKSKEDDLET--AEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSAD 352
Db    447 KRNEEVVRSSEEENIESLAAFKKESQHPLGCPEEEIQRVERLIEKE------GQESLSSP 500

Qy    353 YEKVELPLEDQVGD--LEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQG 410
Db    501 EE------EDQETDRPLEKENGEPLKPVEEE--DQLFETLIEKEGQESLSSPEEEDQETD 552

Qy    411 GGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAE-----------PAEELMKSREMCVS 458
Db    553 RPLEKEEDQLVERLVEKEGQESLSSPEEEDQETDRPLEKENGEPLKPVEEEDQLFETLIE 612
```

```
Qy    459 GGDHTQLTDLSPEE------KTLPKHPEGIVSEV------EMLSS-QERIKVQGSPLKK- 504
                |:  |||      :  |  |   : :|   :        | ||| :|   :     ||:|
Db    613 KEGQESLS--SPEEEDQETDRPLEKEEDQLVERLVEKEGQESLSSPEEEDQETDRPLEKE 670

Qy    505 -------LFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSP 557
                 |     | :  ||  ::: :        : | |  :     :    |   |::|:  | |||
Db    671 EDQLVETLIEKQGQECLSSPEEEDQE----TDRPLEKE--EDQRVERLIEKEGQESLSSP 724

Qy    558 EEPEETT--CLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSE 615
          || :: |    |||       ::||                          |  :      :|| |  |
Db    725 EEEDQETYRLLEK------ENGE-------------------PLKPVEEEDQRVERLIE 758

Qy    616 SDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSS 675
            : :|  |   :|                 | |   |||:  |::   |  | :|  ::   : :: : |
Db    759 KEDQESLSSLK-----------EEDQRIVKTL--ERENWESLRSLDENL--DTVMPLESK 803

Qy    676 KKRARKASSSDDE------------GGPRTLGGDSHRAEEASKDKEAGTDAVPASTQE-- 721
            ::  |:  :::|                      :|   ::        : :|  :       :    |||
Db    804 NQKPLKSLEAEEEQRIVKPLEKVSQDSIGSLEKENVELLRSLEDDQITESLLKKGTQESL 863

Qy    722 ---QDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKL----EEKAEDSSVEQ 774
             :|:  |  :   |:          ||:|:     |:|       |:::   |       |||       |:|:
Db    864 ESHEDRNQETQDPQ-RFLEEEGQGI-----VKQL---EKENQGFLGSLEEEKVVKRSLER 914

Qy    775 LSTEIEPSREESWVSIK----------KFIPGRRKKRADGKQEQATVEDSGPVEINEDDP 824
           :  |     |  |: ||:                |   :  |:     |   ||::   ::
Db    915 ENHEPLSSVEKDWVTESLLERESQDSGKSLEGQEAFRCLGKEDPESL------------ 961

Qy    825 NVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAV-----------YVSEELSKTLVHTVS 873
                        ::   |:  ::::        |    |  |||:             |  || |  |
Db    962 ---------QFPEVQDQEIQRSLQQETQQTLGALGDEQMASEPPEKVGPELLKFLG---- 1008

Qy    874 VAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEA-MPPVEEVTEKDIIAEETPVLTQT 932
              :|     | |:||::     :   || |      ||   :  |    :| :     |       |:   ::
Db   1009 ----NGQEIVRSLEEQNQESL-VSVKETSVETAKSSDMEDIEPLKSAD---EDLEIINSA 1060

Qy    933 LPEGKDAHDDMVTSEVDFTSEAVTATETSEAL----RTEEVTE--ASGAEETTDMVSAVS 986
                  | :  :  :  ||      :            |    |:|          :|:              | ||      :
Db   1061 -----GAQESLWSMEVTRETTRPLEKEVQESLGFVGGNQEILRPLERGNEELGSLGKWNL 1115

Qy    987 QLTDSPDTTEEA-------------------TPVQEVESGVLDTEEEER-QTQAILQAVA 1026
           :   |||:   ||                                 :  |||  |:    :    ::: :  :|:   |  |
Db   1116 EAVDSPEAVEEGRQPLGEEACLEMGEHQEPPRSLGEVEQGLPGSGNQQKWEDRAVENAAA 1175

Qy   1027 DKVKEESQVPATQTVQRTGSKALEKVE-------EVEEDSE------------VLASEK 1066
          |          |  |      |:  |||  :: :: |             | |||:|                  ||| :
Db   1176 D------QGP---TLGRTGIESEDEAELPLSGQGEKEEDAEERELQLDAMGEAWSLASSE 1226

Qy   1067 EKDVMPKGPVQEAGAEHLAQGSSE------------TGQATPESLEVPEVTADV------- 1107
            ::   |:  |   |  |:       ||  |                      ||   ||   |  |   :   ||
Db   1227 PQE--PRVP-SEGGSAGAPQGLEGQSEQVGVLGVPVGQGMPEVTE-PLLQEDVAQAGKRD 1282

Qy   1108 -------DHVATCQVIKLQQ-----------LMEQAVAPESSETLT------DSETNGST 1143
                        ||     :  |:|                    |:|:  |        |        |  |    |
Db   1283 PIELTLGSDAATRAGLGLEQEVAGSGGSRHLAREEAIHPSLGEESVEAKIAHDLEGPGKE 1342

Qy   1144 PLADSDTADGTQQDETIDSQDSKATAAVRQSQ-----VTEEEAATAQKEEPSTLPNNVPA 1198
          |  ::     :       :            |   |       :  :||:                ||  :   |    |    |    |:
Db   1343 P-KEAGALESEISELPRTSSDVLESKGCKQSEPVLGWAVEEASVEASDHEGSDAPEPRPS 1401

Qy   1199 QEEHGEE--------PGRDVLEPTQQELTAAAVPVL-----------AKTEVGQEGEVDW 1239
          :  |  |:|                  ||    ::||          :    |:|                    |  |:: |
Db   1402 ETE-GDEGAQAALAPPGPKLMEP------CSPTPILKDAYEWQPQAEGTRETGRQLEGGS 1454

Qy   1240 LDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDV 1299
                |:|::|||                                       |:  |                     :  ||      |
Db   1455 ASLERVEDEQE--------------------FGLGG-------------IPEGLQ--DW 1478

Qy   1300 EKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQD 1359
          |: :  |:: ::: |                                               :| |        ||       ||: |  |
Db   1479 EESREESEADELGE---------------------TLPDS---TPLGLYLRSPTSPKWD 1513

Qy   1360 KAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWT 1419
            :||            :   ::      ||                          :  |    |       |   |                 || :
Db   1514 QAG---------EQRLSPQGEA----------RKEGWGPAVPAAQGLSNPPGEEERGHDS 1554

Qy   1420 LQHAEDTVPLGPESQAESIPII----------VTPAPESTLHPDL--QG---------EI 1458
            :|:         ||  |   |   :|  :                   |  |  ||        |:    ||          |
Db   1555 DLSSEEFEDLGTE--ASLLPGLPKEVADHLGQVPPGPE----PECWDQGGESDGFADEEE 1608

Qy   1459 SASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELES 1501
             |   :   |   ||   ||    ::|       |      :  ::  |  :       ::  |  ||
Db   1609 SGEEGEEEEHEDGTESGAQWWGSGPSKVQHVTQ-RGDLQEHES 1650
```

Search completed: December 13, 2002, 00:16:36
Job time : 93 secs

```
OM protein - protein search, using sw model

Run on:         December 12, 2002, 21:09:06 ; Search time 49 Seconds
                                          (without alignments)
                                          1350.943 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters:      112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SwissProt_40:*

      Pred. No. is the number of results predicted by chance to have a
      score greater than or equal to the score of the result being printed,
      and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4294	53.2	1781	1	AKAC_HUMAN	Q02952 homo sapien
2	501	6.2	2464	1	MAPB_MOUSE	P14873 mus musculu
3	493	6.1	3924	1	ANK2_HUMAN	Q01484 homo sapien
4	492	6.1	2468	1	MAPB_HUMAN	P46821 homo sapien
5	470	5.8	2459	1	MAPB_RAT	P15205 rattus norv
6	433.5	5.4	2774	1	MAPA_RAT	P34926 rattus norv
7	421	5.2	2805	1	MAPA_HUMAN	P78559 homo sapien
8	411.5	5.1	3562	1	PGCV_CHICK	Q90953 gallus gall
9	408.5	5.1	4687	1	PLE1_RAT	P30427 rattus norv
10	402.5	5.0	4473	1	PLE1_CRIGR	Q9ji55 cricetulus
11	400.5	5.0	4684	1	PLE1_HUMAN	Q15149 homo sapien
12	395	4.9	3396	1	PGCV_HUMAN	P13611 homo sapien
13	388	4.8	1805	1	NEST_RAT	P21263 rattus norv
14	388	4.8	1861	1	MAP2_RAT	P15146 rattus norv
15	385.5	4.8	1087	1	NFH_MOUSE	P19246 mus musculu
16	382	4.7	2116	1	MYS2_DICDI	P08799 dictyosteli
17	379	4.7	2349	1	TPR_HUMAN	P12270 homo sapien
18	377.5	4.7	1827	1	MAP2_HUMAN	P11137 homo sapien
19	376	4.7	3256	1	KI67_HUMAN	P46013 homo sapien
20	374	4.6	1233	1	YFI6_YEAST	P43597 saccharomyc
21	374	4.6	1637	1	MRSP_STAAU	P80544 staphylococ
22	371.5	4.6	1828	1	MAP2_MOUSE	P20357 mus musculu
23	371.5	4.6	3911	1	AKA9_HUMAN	Q99996 h a-kinase
24	371	4.6	1658	1	YM67_YEAST	Q03661 saccharomyc
25	370	4.6	1391	1	MST2_DROHY	Q08696 drosophila
26	369.5	4.6	3381	1	PGCV_BOVIN	P81282 bos taurus
27	367.5	4.6	1411	1	TCOF_HUMAN	Q13428 homo sapien
28	367	4.5	1020	1	NFH_HUMAN	P12036 homo sapien
29	366.5	4.5	1972	1	P53I_HUMAN	Q12888 homo sapien
30	363	4.5	831	1	NFH_RAT	P16884 rattus norv
31	363	4.5	2453	1	NCRI_MOUSE	Q60974 mus musculu
32	354	4.4	1618	1	NEST_HUMAN	P48681 homo sapien
33	352.5	4.4	1616	1	P200_MYCGE	Q49429 mycoplasma
34	346.5	4.3	2017	1	MYSN_DROME	Q99323 drosophila
35	346.5	4.3	2230	1	GOG4_HUMAN	Q13439 homo sapien
36	339.5	4.2	3358	1	PGCV_MOUSE	Q62059 mus musculu
37	338.5	4.2	2492	1	ATRX_HUMAN	P46100 homo sapien
38	337.5	4.2	4377	1	ANK3_HUMAN	Q12955 homo sapien
39	335.5	4.2	1220	1	IF2P_HUMAN	O60841 homo sapien
40	335	4.1	1978	1	MYHB_CHICK	P10587 gallus gall
41	334.5	4.1	1395	1	SP41_YEAST	P38904 saccharomyc
42	334	4.1	1744	1	TANA_XENLA	Q01550 xenopus lae
43	332.5	4.1	1676	1	APSA_EMENI	Q00083 emericella
44	332	4.1	1972	1	MYHB_MOUSE	O08638 mus musculu
45	331	4.1	1938	1	MYS_AEQIR	P24733 aequipecten

ALIGNMENTS

```
RESULT 1
AKAC_HUMAN
ID   AKAC_HUMAN     STANDARD;      PRT;  1781 AA.
AC   Q02952; Q99970; O00498; O00310;
DT   01-FEB-1994 (Rel. 28, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   A-kinase anchor protein 12 (A-kinase anchor protein 250 kDa) (AKAP
DE   250) (Myasthenia gravis autoantigen gravin).
GN   AKAP12 OR AKAP250.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Heart;
RX   MEDLINE=97153077; PubMed=9000000;
RA   Nauert J.B., Klauck T.M., Langeberg L.K., Scott J.D.;
RT   "Gravin, an autoantigen recognized by serum from myasthenia gravis
RT   patients, is a kinase scaffold protein.";
RL   Curr. Biol. 7:52-62(1997).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=98269042; PubMed=9604001;
RA   Sato N., Kokame K., Shimokado K., Kato H., Miyata T.;
RT   "Changes of gene expression by lysophosphatidylcholine in vascular
RT   endothelial cells: 12 up-regulated distinct genes including 5 cell
RT   growth-related, 3 thrombosis-related, and 4 others.";
RL   J. Biochem. 123:1119-1126(1998).
RN   [3]
RP   SEQUENCE OF 43-1781 FROM N.A.
RC   TISSUE=Umbilical vein endothelial cells;
RA   Bowditch R.D., Ginsberg M.H.;
RT   "Sequence of gravin cDNA isolated from a human umbilical vein
RT   endothelial cell library.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 1477-1781 FROM N.A.
RC   TISSUE=Umbilical vein endothelial cells;
RX   MEDLINE=92395179; PubMed=1522245;
RA   Gordon T., Grove B., Loftus J.C., O'Toole T., McMillan R.,
RA   Lindstrom J., Ginsberg M.H.;
RT   "Molecular cloning and preliminary characterization of a novel
RT   cytoplasmic antigen recognized by myasthenia gravis sera.";
RL   J. Clin. Invest. 90:992-999(1992).
CC   -!- FUNCTION: ANCHORING PROTEIN THAT MEDIATES THE SUBCELLULAR
CC       COMPARTMENTATION OF PROTEIN KINASE (PKA) AND PROTEIN KINASE C
CC       (PKC).
CC   -!- SUBUNIT: BINDS TO DIMERIC RII-ALPHA REGULATORY SUBUNIT OF PKA.
CC   -!- SUBCELLULAR LOCATION: CYTOPLASMIC. MAY BE PART OF THE CORTICAL
CC       CYTOSKELETON.
CC   -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; MAY BE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: EXPRESSED IN ENDOTHELIAL CELLS, CULTURED
```

```
CC       FIBROBLASTS AND OSTEOSARCOMA, BUT NOT IN PLATELETS, LEUKOCYTES,
CC       MONOCYTIC CELL LINES OR PERIPHERICAL BLOOD CELLS.
CC   -!- INDUCTION: ACTIVATED BY LYSOPHOSPHATIDYLCHOLINE (LYSOPC).
CC   -!- DOMAIN: POLYBASIC REGIONS LOCATED BETWEEN RESIDUES 265 AND 556 ARE
CC       INVOLVED IN BINDING PKC.
CC   -!- DISEASE: ANTIBODIES TO THE C-TERMINAL OF GRAVIN CAN BE PRODUCED BY
CC       PATIENTS WITH MYASTHENIA GRAVIS (MG).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U81607; AAC51366.1; -.
DR   EMBL; AF001504; AAB58938.1; -.
DR   EMBL; AB003476; BAA19927.1; -.
DR   EMBL; M96322; AAA35931.1; -.
DR   PIR; A43922; A43922.
DR   Genew; HGNC:370; AKAP12.
DR   MIM; 604698; -.
DR   InterPro; IPR001573; PkinA_anch.
KW   Antigen; Repeat; Alternative splicing.
FT   DOMAIN      603    633       AKAP 1.
FT   DOMAIN      752    782       AKAP 2.
FT   DOMAIN      797    827       AKAP 3.
FT   DOMAIN       98    101       POLY-GLU.
FT   DOMAIN      265    556       INVOLVED IN PKC-BINDING (PROBABLE).
FT   DOMAIN     1540   1553       RII-BINDING (PROBABLE).
FT   VARSPLIC      1     98       MISSING (IN ISOFORM 2).
FT   VARSPLIC     99    106       EEEVIVTE -> MLGTITIT (IN ISOFORM 2).
FT   CONFLICT    117    117       E -> K (IN REF. 2).
FT   CONFLICT    142    144       NRN -> TPEI (IN REF. 2 AND 3).
FT   CONFLICT    215    215       Q -> K (IN REF. 2).
FT   CONFLICT    448    448       G -> E (IN REF. 2 AND 3).
FT   CONFLICT    694    694       R -> G (IN REF. 2 AND 3).
FT   CONFLICT    867    867       G -> S (IN REF. 2 AND 3).
FT   CONFLICT    986    986       S -> A (IN REF. 2 AND 3).
FT   CONFLICT   1530   1530       E -> EE (IN REF. 3 AND 4).
FT   CONFLICT   1581   1581       V -> M (IN REF. 4).
FT   CONFLICT   1601   1601       Q -> L (IN REF. 2).
SQ   SEQUENCE   1781 AA;  191439 MW;  BA813937379FAC0F CRC64;

  Query Match             53.2%;  Score 4294;  DB 1;  Length 1781;
  Best Local Similarity   57.9%;  Pred. No. 7.8e-140;
  Matches  959;  Conservative  187;  Mismatches  401;  Indels  108;  Gaps   35;

Qy          1 MGAGSSTEQRSPEQ-PAGSDTPSELVLSGHGPAAEAS-GAAGDPADA--DPATKLPQKNG 56
Db          1 MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADPAIAASDPATKLLQKNG 60

Qy         57 QLSSVNGVAEQGDVHVQEENQEGQ--------------EEEVVDEDVGQRESEDVREKDR 102
Db         61 QLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDS 120

Qy        103 VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVG 162
Db        121 DKEMATKSAVVHDITDDGQEENRN-IEQIPSSESNLEELTQPTESQANDIGFKKVFKFVG 179

Qy        163 FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQS 222
Db        180 FKFTVKKDKTEKPDTVQLLTVKKDEGEGA---AGAGDHQDPSL--GAGEAASKESEPKQS 234

Qy        223 TEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTS 282
Db        235 TEKPEETLKREQSHAEISPPAESGQAV-EECKEEGEEKQEKEPSKSAESPTSPVTSETGS 293

Qy        283 SFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEE----- 337
Db        294 TFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASE 353

Qy        338 ----QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQ-E 392
Db        354 QAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEE 413

Qy        393 VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db        414 VVAEVHVSTVEERTEEQ--------KTEVEETAGSVPAEELVGMDAEPQEAEPAKELVKL 465

Qy        453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db        466 KETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLK 525

Qy        513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLE 572
Db        526 KLSGKKQKGKR-GGGDEESGEHTQVPADSPDSQEEQKGESSASSPEEPEEITCLEKGLAE 584

Qy        573 APQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSS 632
Db        585 VQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSS 644

Qy        633 TDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPR 692
Db        645 TESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALICVGSSKKRARRSSSDEEGGPK 704

Qy        693 TLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRL 752
Db        705 AMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRL 764

Qy        753 VTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQ 808
Db        765 VTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQ 824

Qy        809 ATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELS 865
Db        825 APVEDAGPTGANEDDSDVPAVVPLSEYDAVEREKMEAQQAQKGAEQPEQKAATEVSKELS 884

Qy        866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA-E 924
Db        885 ESQVHMMAAAVADGTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEE 944

Qy        925 ETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSA 984
Db        945 EPPTVTEPLPENREARGDTVVSEAELTPEAVTAAETAGPLGSEEGTEASAAEETTEMVSA 1004

Qy        985 VSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPAT------ 1038
Db       1005 VSQLTDSPDTTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDV 1064

Qy       1039 -QTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPES 1097
Db       1065 LQPVQRA------EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPES 1118

Qy       1098 LE-VPEVTADVDH---VATCQV-----IKLQQL-MEQAVAPESSETLTDSETNGSTPLAD 1147
Db       1119 FEKAPQVTESIESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVAD 1178

Qy       1148 SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPG 1207
Db       1179 FDAPGTTQKDEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPA-PSSFVFQEETKEQSK 1237

Qy       1208 -RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVK-----EEQEVFVHSG--PNS 1259
Db       1238 MEDTLEHTDKEVSVETVSILSKTEGTQ--EADQYADEKTKDVPFFEGLEGSIDTGITVSR 1295

Qy       1260 QKAADVTYDSEVMGVAGCQEKESTEVQS-----LSLEEGEMETDVEKEKRETKPEQVSEE 1314
Db       1296 EKVTEVALKGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEE 1355

Qy       1315 G-EQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS-PSLPDQDKAGCIEVQVQSLD 1372
Db       1356 KLEHETAVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSE 1415

Qy       1373 TTVTQTAEAVEKVI--ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
```

```
                 : | || | |: :  ||  | ||||: |  ||||: ||||       :    || || |
Db     1416 ASFTLTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTG 1475

Qy     1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
              |: ||:| |:||:    :  |  ||:|| : | : :|:| |:  |  :   | | |||  :
Db     1476 PDCQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEV--KVSVAIEDL- 1532

Qy     1491 KAEPE--ILELESKSNKIVLNVIQTAVDQFART-ETAPETHAYDSQTQVPACRLDSREPN 1547
                |||  |||||:||:|:| |:||||||| || ||| |      : |||     : ||::
Db     1533 --EPENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAG 1590

Qy     1548 RCWTK-----MKDAKMKHPVPQPREDLQVLTVLEA 1577
              :   |           |: : |:   :|: :   | :|
Db     1591 QETEKEGEEPQASAQDETPITSAKEESESTAVGQA 1625


RESULT 2
MAPB_MOUSE
ID   MAPB_MOUSE     STANDARD;      PRT;  2464 AA.
AC   P14873;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) (MAP1.2) (MAP1(X))
DE   [Contains: MAP1 light chain LC1].
GN   MAP1B OR MTAP1B OR MTAP5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND DOMAIN.
RC   STRAIN=Swiss Webster; TISSUE=Brain;
RX   MEDLINE=90094539; PubMed=2480963;
RA   Noble M., Lewis S.A., Cowan N.J.;
RT   "The microtubule binding domain of microtubule-associated protein
RT   MAP1B contains a repeated sequence motif unrelated to that of MAP2
RT   and tau.";
RL   J. Cell Biol. 109:3367-3376(1989).
CC   -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC       PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC       THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC       TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC       MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC       STABILIZING MICROTUBULES.
CC   -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC       WITH MAP1A AND MAP1B PROTEINS.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 IS COEXPRESSED WITH MAP1B. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1B BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B. IT INTERACTS WITH THE AMINO-TERMINAL REGION
CC       OF MAP1B.
CC   -!- SIMILARITY: TO MAP1A.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X51396; CAA35761.1; -.
DR   PIR; S07549; QRMSP1.
DR   MGD; MGI:1306778; Mtap1b.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 7.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN           ?   2464       MAP1 LIGHT CHAIN LC1.
FT   REPEAT       1874   1890       MAP1B 1.
FT   REPEAT       1891   1907       MAP1B 2.
FT   REPEAT       1908   1924       MAP1B 3.
FT   REPEAT       1925   1941       MAP1B 4.
FT   REPEAT       1942   1958       MAP1B 5.
FT   REPEAT       1959   1975       MAP1B 6.
FT   REPEAT       1993   2009       MAP1B 7.
FT   REPEAT       2010   2026       MAP1B 8.
FT   REPEAT       2027   2043       MAP1B 9.
FT   REPEAT       2044   2060       MAP1B 10.
FT   DOMAIN        589    787       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                  KKEE AND KKEI/V REPEATS).
SQ   SEQUENCE   2464 AA;  270408 MW;  FBD3DD99CFDBDA87 CRC64;

  Query Match             6.2%;  Score 501;  DB 1;  Length 2464;
  Best Local Similarity   21.1%;  Pred. No. 2.4e-10;
  Matches  387;  Conservative  266;  Mismatches  648;  Indels  532;  Gaps   86;

Qy       45 ADPATKL-----PQKNGQLSSVNGVAEQGDVH------VQEENQEGQEEEVVDEDVGQRE 93
            |:|| |:      |  : | : : |: :   :        ::: ||       :  |  ::
Db      478 ANPAEKIIRVLFPGNSTQYNILEGLEKLKHLDFLKQPLATQKDLTGQVPTPPVKQVKLKQ 537

Qy       94 SEDVREKDRVEEMAANSTAVEDITKDGQEETSEI-----IEQIPASENNVEEMVQPAESQ 148
              | ||     : |    | : : |: :||| |:      :|: |  | |:
Db      538 RADSRES---LKPATKPVASKSVRKESKEETPEVTKTSQVEKTPKVESKE---------- 584

Qy      149 ANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETA 208
                           |  |||||  |:::    :|           |  | : : | | |:
Db      585 --------------KVLVKKDKPVKTESKPSVT---------EKEVSSKEEQSP-VKAE 619

Qy      209 VGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKS 268
             | |  : ||: | : :|        :|:|         :: : : ||: | |  :|::| | |
Db      620 VAEKQATESKPKVTKDK---VVKKE-------IKTKLEEKKEEKPKKEVVKKEDKTPLKK 669

Qy      269 PESP-SSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEE 327
             | |  |   |      ||           ||:    :  ||  |: |:|  ::| :| |
Db      670 DEKPRKEEVKKEIKKEIKK-------EERKELKKEVKKETPLKDAKKEVKKEEKK--EVK 720

Qy      328 KEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDE-- 385
            ||: ||  | :: ::|  ::    :| :|     |: :|    | |:::: |||      :
Db      721 KEEKEPKKEIKKISKDIKKSTPQSDTKKPS-ALKPKVAKKEESTKKE--PLAAGKLKDKG 777

Qy      386 -----KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVP 440
                  ||  |     |  |  |       |  |  |  |  |    |        :: |:::
Db      778 KVKVIKKEGKTTEAAATAVGTAATTAAVVAAAGIAASGPVKELEAER---SLMSSPEDLT 834

Qy      441 QEAEPAEELMKSREMCVSGGDHTQLTDLSPEEK---TLPKHPEGIVSEVEMLSSQERIKV 497
            :: :     :|: |: |:        ||  :  |||  ||        |  | |:
Db      835 KDFEE----LKAEEIDVAKDIKPQLELIEDEEKLKETQPGEAYVIQKETEVSKGSAE--- 887

Qy      498 QGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSP 557
              ||          |:        |:        |   ::  |  | :  :  :  :  | :| |
Db      888 --SP------DEGITTTEGE-------GECEQTPEELEPVEKQGVD--DIEKFEDEGAGF 930

Qy      558 EEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESD 617
            || ||       |   |    :  |   |   |||      |          |     |:| |
Db      931 EESSET-----GDYEEKAETEEAEEPEEDGEDNASG-----SASK--------HSPTEDD 972

Qy      618 KEEELE-----KVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICV 672
            :  :  |     | |  :: |  |     |  |:|   || :: ||        |
Db      973 ESAKAEADVHLKEKRESVVSGDDRAEEDMDDVLEKGEAEQSEEEGEEED----------- 1021

Qy      673 GSSKKRARKASSSDDEG-GPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSS-- 729
                    ||  : :||  |        : :     |  |   |||        |:||    |:|
Db     1022 --------KAEDAREEGYEPDKTEAEDYVMAVADKAAEAGV------TEEQYGYLGTSAK 1067

Qy      730 -------SPEPAGS--------PSEGEGVSTWESFK-----RLVTPRKKSKSKLEEKAED 769
                   | ||| |            || |    : : |              ::| :|  :|
Db     1068 QPGIQSPSREPASSIHDETLPGGSESEATASDEENREDQPEEFTATSGYTQSTIEISSEP 1127

Qy      770 SSVEQLST----------EIEPSREESWVSIKKF--------------------IPGRR 798
            : ::::||          |   |  : :|:| |:                      | :
Db     1128 TPMDEMSTPRDVMSDETNNEETESPSQEFVNITKYESSLYSQEYSKPAVASFNGLSEGSK 1187
```

```
Qy     799 KKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAV 858
               |||   |:     |     |:|    :    | :    | :::|     ::
Db    1188 TDATDGKDYNASASTISPPSSMEEDKFSKSA--LRDAYCSEEKELKASAELDIKD----- 1240

Qy     859 YVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHT-AGE-----AMPP 912
            ||:|         :| |              :||| :| |   |:|  |  ||      :: |
Db    1241 -VSDE-------RLSPA-------------KSPS-LSPSPPSPIEKTPLGERSVNFSLTP 1278

Qy     913 VE-EVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTAT-----------E 959
            | :|: :         :| :|| : |   |    :   | ||   |::||  :         |
Db    1279 NEIKVSAEGEARSVSPGVTQAVVEEHCASPEEKTLEVVSPSQSVTGSAGHTPYYQSPTDE 1338

Qy     960 TSEALRTE------------EVTEASGAEETT--------------------------- 979
            |  | ||              | :||    |
Db    1339 KSSHLPTEVSENAQAVPVSFEFSEAKDENERASLSPMDEPVPDSESPVEKVLSPLRSPPL 1398

Qy     980 --------DMVSAVSQL----TDSP--------DTTEEATPVQEVES-GVLDTEEEERQT 1018
                    | :|| |::     ::||               :  :|| :: | |:    ::||: |
Db    1399 LGSESPYEDFLSADSKVLGRRSESPFEGKNGKQGFPDRESPVSDLTSTGLYQDKQEEKST 1458

Qy    1019 QAILQAVADKVKEE-SQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQ 1077
              |         :||:      |   |:     |::    ::| :    :|   |   : ||    |   :
Db    1459 GFI------PIKEDFGPEKKTSDVETMSSQSALALDERKLGGDV--SPTQIDVSQFGSFK 1510

Qy    1078 EAGAEHLAQGSETGQ-ATPESLEVPEVT-ADVDHVATCQVIKLQQLMEQAVAPESSETLT 1135
            |      :::|: :  : |||      |  |  |  : ::  ||:            :  |   :: :
Db    1511 EDTKMSISEGTVSDKSATPVDEGVAEDTYSHMEGVASV-----------STASVATSSFP 1559

Qy    1136 DSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEE------- 1188
            :  |:  :|    ::     |:     :|   ||  : :  |:      :|    :   |||
Db    1560 EPTTDDVSPSLHAEV--GSPHSTEVD--DSLSVSVVQTPTTFQETEMSPSKEECPRPMSI 1615

Qy    1189 ------PSTLPNNVPAQEEHGEEPGRDV-----------------LEPTQQELTAAAVPV 1225
                   | |   :   | |:    |:      :                      |      | |:   |
Db    1616 SPPDFSPKTAKSRTPVQDHRSEQSSMSIEFGQESPEHSFAMDFSRQSPDHPTLGAS---V 1672

Qy    1226 LAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYD---SEVMGVAGCQEKES 1282
            |  ||  |     |||:           :     |   |  ::: :   |  |      ||:: |:  :     |
Db    1673 LHITENGPT-EVDY---SPCDIQDSSLSHKIPPTEEPS-YTQDNDLSELISVSQVEASPS 1727

Qy    1283 T-----------------------------------EVQSLSLEEGEMETDVEK-EKRE 1305
            |                                      :||||   |:     ::|:     ||
Db    1728 TSSAHTPSQIASPLQEDTLSDVVPPREMSLYASLASEKVQSLEGEKLSPKSDISPLTPRE 1787

Qy    1306 TKP-----EQVSEEGEQETAAPEHEGTYGKPV-----------------------LTLD 1336
            : |         |      :||||   |: :    |:                           | |
Db    1788 SSPLYSPGFSDSTSAAKETAA-AHQASSSPPIDAATAEPYGFRSSMLFDTMQHHLALNRD 1846

Qy    1337 MPSS-----ERGKALG---------SLGGSPSLPDQDKAGCIEVQVQSLDT--TVTQTA 1379
            : :|       ||   |             :   |||    ||     |   |  ::: |     |
Db    1847 LTTSSVEKDSGGKTPGDFNYAYQKPENAAGSPDEEDYD----YESQEKTIRTHDVVRYYY 1902

Qy    1380 EAVEKVIETVVISETGESPECVGAHLLPAEKSSAT--GGHWTLQHAEDTVPLGPESQAES 1437
            |   |: |::      ::| | | :              ||:: |    | :| :  | |       ||      |
Db    1903 EKTERTIKSP--CDSGYSYETI-------EKTTKTPEDGGYTCEITEKTTRT-PEEGGYS 1952

Qy    1438 IPIIVTPAPESTLHPDLQGEISASQRERSEE---------EDKPDAGPD-ADGKESTAIE 1487
               |      : ::|   |::  |    :   :  |||                 :|   |   |       |
Db    1953 YEI----SEKTTRTPEVSG-YTYEKTERSRRLLDDISNGYDDTEDGGHTLGDCSYSYETT 2007

Qy    1488 KVLKAEPEILELESKSNKIVLNVIQTAVDQFARTETAPETHAY-----DSQTQVPACRLD 1542
            :  :  :   |       | ||  |                :     |    :|:|  ||          :    |:  |
Db    2008 EKITSFP---ESESYS-----------YETSTKTTRSPDTSAYCYETMEKITKTPQASTY 2053

Qy    1543 SRE-PNRCWTKMKDAKMKHPVPQPREDLQVLTV 1574
            | |    :||:|   |       | |   : |:|: :   |
Db    2054 SYETSDRCYTTEK----KSP-SEARQDVDLCLV 2081
```

RESULT 3
ANK2_HUMAN

```
ID   ANK2_HUMAN     STANDARD;      PRT;  3924 AA.
AC   Q01484; Q01485;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Ankyrin 2 (Brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).
GN   ANK2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC   TISSUE=Brain stem;
RX   MEDLINE=91302466; PubMed=1830053;
RA   Otto E., Kunimoto M., McLaughlin T., Bennett V.;
RT   "Isolation and characterization of cDNAs encoding human brain
RT   ankyrins reveal a family of alternatively spliced genes.";
RL   J. Cell Biol. 114:241-253(1991).
RN   [2]
RP   REVISIONS.
RA   Carpenter S.;
RL   Submitted (MAY-1999) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain stem;
RX   MEDLINE=94075409; PubMed=8253844;
RA   Chan W., Kordeli E., Bennett V.;
RT   "440-kD ankyrinB: structure of the major developmentally regulated
RT   domain and selective localization in unmyelinated axons.";
RL   J. Cell Biol. 123:1463-1473(1993).
RN   [4]
RP   SEQUENCE OF 463-495 FROM N.A.
RX   MEDLINE=92009921; PubMed=1833308;
RA   Tse W.T., Menninger J.C., Yang-Feng T.L., Francke U., Sahr K.E.,
RA   Lux S.E., Ward D.C., Forget B.G.;
RT   "Isolation and chromosomal localization of a novel nonerythroid
RT   ankyrin gene.";
RL   Genomics 10:858-866(1991).
CC   -!- FUNCTION: Attach integral membrane proteins to cytoskeletal
CC       elements. Also bind to cytoskeletal proteins.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
CC       CELLS THROUGHOUT THE BRAIN.
CC   -!- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
CC       AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
CC       AND FUNCTION (POTENTIAL).
CC   -!- SIMILARITY: CONTAINS 23 ANK REPEATS.
CC   -!- SIMILARITY: CONTAINS 1 DEATH DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X56957; CAA40278.1; -.
DR   EMBL; X56958; CAA40279.2; -.
DR   EMBL; Z26634; CAB42644.1; -.
DR   EMBL; M37123; AAA62828.1; -.
DR   PIR; S14533; S14533.
DR   PIR; A39643; A39643.
DR   PIR; B39643; B39643.
DR   PIR; S14569; S14569.
DR   HSSP; P42771; 1DC2.
DR   Genew; HGNC:493; ANK2.
DR   MIM; 106410; -.
DR   InterPro; IPR002110; ANK.
DR   InterPro; IPR000488; Death.
DR   InterPro; IPR000906; ZU5.
DR   Pfam; PF00023; ank; 24.
DR   Pfam; PF00531; death; 1.
```

```
DR   Pfam; PF00791; ZU5; 1.
DR   PRINTS; PR01415; ANKYRIN.
DR   SMART; SM00248; ANK; 21.
DR   SMART; SM00005; DEATH; 1.
DR   SMART; SM00218; ZU5; 1.
DR   PROSITE; PS50088; ANK_REPEAT; 20.
DR   PROSITE; PS50297; ANK_REP_REGION; 1.
DR   PROSITE; PS50017; DEATH_DOMAIN; 1.
KW   Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
KW   Phosphorylation.
FT   REPEAT       63     92       ANK 1.
FT   REPEAT       96    125       ANK 2.
FT   REPEAT      129    158       ANK 3.
FT   REPEAT      162    191       ANK 4.
FT   REPEAT      193    220       ANK 5.
FT   REPEAT      232    261       ANK 6.
FT   REPEAT      265    294       ANK 7.
FT   REPEAT      298    327       ANK 8.
FT   REPEAT      331    360       ANK 9.
FT   REPEAT      364    393       ANK 10.
FT   REPEAT      397    426       ANK 11.
FT   REPEAT      430    459       ANK 12.
FT   REPEAT      463    492       ANK 13.
FT   REPEAT      496    525       ANK 14.
FT   REPEAT      529    558       ANK 15.
FT   REPEAT      562    591       ANK 16.
FT   REPEAT      595    624       ANK 17.
FT   REPEAT      628    657       ANK 18.
FT   REPEAT      661    690       ANK 19.
FT   REPEAT      694    723       ANK 20.
FT   REPEAT      727    756       ANK 21.
FT   REPEAT      760    789       ANK 22.
FT   REPEAT      793    822       ANK 23.
FT   DOMAIN     1773   1950       REPEAT-RICH REGION.
FT   REPEAT     1773   1784       REPEAT A.
FT   REPEAT     1785   1796       REPEAT A.
FT   REPEAT     1797   1808       REPEAT A.
FT   REPEAT     1809   1820       REPEAT A.
FT   REPEAT     1821   1832       REPEAT A.
FT   REPEAT     1833   1844       REPEAT A.
FT   REPEAT     1845   1856       REPEAT A.
FT   REPEAT     1857   1867       REPEAT A (APPROXIMATE).
FT   REPEAT     1868   1879       REPEAT A.
FT   REPEAT     1880   1891       REPEAT A.
FT   REPEAT     1892   1902       REPEAT A (APPROXIMATE).
FT   REPEAT     1903   1914       REPEAT A.
FT   REPEAT     1915   1926       REPEAT A.
FT   REPEAT     1927   1938       REPEAT A.
FT   REPEAT     1939   1950       REPEAT A.
FT   DOMAIN     3536   3620       DEATH.
FT   VARSPLIC   1039   1039       Q -> QFLGKLHLPTAPPPLNEGESLVSRILQLGPPGTK
FT                                (IN ISOFORM 2).
FT   VARSPLIC   1444   3528       MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT   CONFLICT    475    476       GQ -> PE (IN REF. 4).
FT   CONFLICT    971    971       I -> S (IN REF. 1).
FT   CONFLICT   3581   3582       QY -> HA (IN REF. 1).
FT   CONFLICT   3586   3586       I -> Y (IN REF. 1).
SQ   SEQUENCE   3924 AA;  430337 MW;  52AC496C428E29D2 CRC64;

  Query Match             6.1%;  Score 493;  DB 1;  Length 3924;
  Best Local Similarity   20.4%;  Pred. No. 7.2e-10;
  Matches  422;  Conservative  270;  Mismatches  680;  Indels  698;  Gaps   99;

Qy         68 GDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSE- 126
                | : |:|  :  :||     :: :|  ||:   ::| |:  : |   |     |   ||
Db       1497 GSIKVKELVKAAEEEPGEPFEIVERVKEDL---EKVNEILRSGTCTRD-ESSVQSSRSER 1552

Qy        127 -IIEQ--IPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDK----NEKSDTV- 178
               ::|:  :  |:  :|| || :  :    :|       | |:||    |  :| :
Db       1553 GLVEEEWVIVSDEEIEEARQKAPLEITEYPCVEVRIDKEIKGKVEKDSTGLVNYLTDDLN 1612

Qy        179 --------QLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKES-----ELKQSTEK 225
                      || ||:   |:  ||                 ||| |: ||       :  |||:|
Db       1613 TCVPLPKEQLQTVQDKAGKKCEA-------------LAVGRSSEKEGKDIPPDETQSTQK 1659

Qy        226 Q-----------EGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPE---- 270
              |           ||::|   |  ||| :::|  |   |    ||    ::|  :||
Db       1660 QHKPSLGIKKPVRRKLKEKQKQKEEGLQASAEKA--ELKKGSSEESLGEDPGLAPEPLPT 1717

Qy        271 -SPSSPVNSET-TSSFKKFFTHGWAGWRKKTSFKKSKEDD------LETAEKRKEQEAEK 322
                :||:  ||   |  |           |  : :|  ||:       |     | ||   :|
Db       1718 VKATSPLIEETPIGSIKD----------KVKALQKRVEDEQKGRSKLPIRVKGKEDVPKK 1767

Qy        323 VDEEEKEKTEPASEEQ-----EPAEDTDQARLSADYEKVE-------------------- 357
                            |: :  :          |:  |::   :   | |
Db       1768 TTHRPHPAASPSLKSERHAPGSPSPKTERHSTLSSSAKTERHPPVSPSSKTEKHSPVSPS 1827

Qy        358 LPLEDQVGDLEASSEEKCAPLA--------TEVFDEKMEAHQEVV--------------- 394
                  |       :|  || :|::          : |   | | |  |
Db       1828 AKTERHSPASSSSKTEKHSPVSPSTKTERHSPVSSTKTERHPPVSPSGKTDKRPPVSPSG 1887

Qy        395 -AEVH-------------VSTVEKTEEEQ--GGGGEAEGGVVVEGTG--ESLPPEKLAEP 436
                | |              ||    :|:: |       |:  |   :  |   :|  |  ||
Db       1888 RTEKHPPVSPGRTEKRLPVSPSGRTDKHQPVSTAGKTEKHLPVSPSGKTEKQPPVSPTSK 1947

Qy        437 QEVPQEAEPAEELMKSREMCVSGGD----------HTQLTDLSPEEK-----TLPKHP-- 479
               |  :|      ||||: :     || |            |             |:||         | |
Db       1948 TERIEETMSVRELMKAFQ---SGQDPSKHKTGLFEHKSAKQKQPQEKGKVRVEKEKGPIL 2004

Qy        480 ---EGIVSEVEMLSSQERIKVQGSPLKK---LFSSSGLKKLSGKKQKGKRGGGGDEE--- 530
                  |    :|  :  :    :|:  | |:    |      || |:||         :    || |:
Db       2005 TQREAQKTENQTIKRGQRLPVTGTAESKRGVRVSSIGVKK--------EDAAGGKEKVLS 2056

Qy        531 ----------PGEYQHIHTESPES--ADEQKGESSASSPEE------------------- 559
                        |   |    :|: | |:    ||||        ||:
Db       2057 HKIPEPVQSVPEEESHRESEVPKEKMADEQGDMDLQISPDRKTSTDFSEVIKQELEDNDK 2116

Qy        560 ------PEETTCLEKGPLEAPQ-----------------------------DGEAE--- 580
                     |||    ||   |    |                              || :|
Db       2117 YQQFRLSEET---EKAQLHLDQVLTSPFNTTFPLDYMKDEFLPALSLQSGALDGSSESLK 2173

Qy        581 ------------EGT---TSDGEKKREGI--TPWASFKKM-VTPKK---RVRRPSESDK 618
                         |||    :|:   | ||:  ||  | : :  :|||    :     || |
Db       2174 NEGVAGSPCGSLMEGTPQISSEESYKHEGLAETPETSPESLSFSPKKSEEQTGETKESTK 2233

Qy        619 EEELEKVKS------------------ATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
               |   :::|                   ||::    | :|   :  |:|   |  |||:
Db       2234 TETTTEIRSEKEHPTTKDITGGSEERGATVTEDSETSTESFQKEATLG-SPKDTSPKRQD 2292

Qy        661 DTSVSWEALICVGSSKKRARKAS-------SSDDEGGPRTLGGDSHRAEEASKDKEAGTD 713
              |        | ||      |  :      ::  || |||  ||  |:|   :  |: :|
Db       2293 D---------CTGSCSVALAKETPTGLTEEAACDE-GQRTFGSSAHKTQTDSEAQE---- 2338

Qy        714 AVPASTQEQDQAQGSSSPEPA-----GSPSEGEGVSTWESFKRLVTPRKKS--------- 759
                    ||  |: :    || :    |: |: :||          |   | : |
Db       2339 ----STATSDETKALPLPEASVKTDTGTESKPQGVIRSPQGLELALPSRDSEVLSAVADD 2394

Qy        760 ------KSKLEEK--AEDSSVEQLSTEIEPS--REESWVSIKKFIPGRRKKRA------- 802
                     | ||    ||:|  :   :|||  :|     :   |   | :|
Db       2395 SLAVSHKDSLEASPVLEDNSSHKTPDSLEPSPLKESPCRDSLESSPVEPKMKAGIFPSHF 2454

Qy        803 -------------------------------DGKQEQATVEDSGPVEINEDDPN----- 825
                                             |  :: : :| ||  :: | |:
Db       2455 PLPAAVAKTELLTEVASVRSRLLRDPDGSAEDDSLEQTSLMESSGKSPLSPDTPSSEEVS 2514

Qy        826 ------------VPAVV-------------------------PLSEYNAVERE--- 841
                          |||:                             :  :: :|:|
Db       2515 YEVTPKTTDVSTPKPAVIHECAEEDDSENGEKKRFTPEEEMFKMVTKIKMFDELEQEAKQ 2574

Qy        842 --------KMEAQGNTELPQLLGAVYVSEELSKTLVHT--------VSVAVIDGTRAVTS 885
                      | |  ::  |    :| | |        ||          | | |    :| |:|
Db       2575 KRDYKKEPKQEESSSSDPDADCSVDVDEP-----KHTGSGEDESGVPVLVTSESRKVSS 2629

Qy        886 VEERSPSW----------------ISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVL 929
                | |                   |        || :        | |||  : ::::
Db       2630 SSESEPELAQLKKGADSGLLPEPVIRVQPPSPLPSSMDSNSSP-EEVQFQPVVSKQ---- 2684
```

```
Qy    930 TQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTE----------------EVTEAS 973
Db   2685 -YTFKMNEDTQEEPGKSEEEKDSESHLA-EDRHAVSTEAEDRSYDKLNRDTDQPKICDGH 2742

Qy    974 GAEETTDMVSA--VSQLTDSP-----------------------DTTEEATPVQEVESG 1007
Db   2743 GCEAMSPSSSARPVSSGLQSPTGDDVDEQPVIYKESLALQGTHEKDTEGEELDVSRAESP 2802

Qy   1008 VLDTEEEERQTQAILQ--AVADKVKEESQVPATQTVQRTG-SKALEKVEEVEED------ 1058
Db   2803 QADCPSESFSSSSSLPHCLVSEGKELDEDISATSSIQKTEVTKTDETFENLPKDCPSQDS 2862

Qy   1059 ---------------SEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATP-------E 1096
Db   2863 SITTQTDRFSMDVPVSDLAENDEIYDPQITSPYENVPSQSFFSSEESKTQTDANHTTSFH 2922

Qy   1097 SLEVPEV--TADVDHVATC----------------QVIKLQQLMEQAVAPESSETLTDSE 1138
Db   2923 SSEVYSVTITSPVEDVVVASSSSGTVLSKESNFEGQDIKMESQLESTL----WEMQSDSV 2978

Qy   1139 TNGSTP-LADSDTADGTQQDETIDSQ---DSKATAAVR------QSQVTEEE----AATA 1184
Db   2979 SSSFEPTMSATTTVVGEQISKVIITKTDVDSDSWSEIREDDEAFEARVKEEEQKIFGLMV 3038

Qy   1185 QKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQEL-TAAAVPVLAKT----EVGQEGEVDW 1239
Db   3039 DRQSQGTTPDTTPAR------------TPTEEGTPTSEQNPFLFQEGKLFEMTRSGAID- 3085

Qy   1240 LDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVM-GVAGCQEKE-STEVQSLSLEEGEMET 1297
Db   3086 MTKRSYADESFHFFQIG---QESREETLSEDVKEGATGADPLPLETSAESLALSESKETV 3142

Qy   1298 DVEKEKRETKPEQVSEEGEQ----------------ETAAPEHEGTYGKPVLTLDMPSSE 1341
Db   3143 D---DEADLLPDSVSEEVEEIPASDAQLNSQMGISASTETPTKEAV---SVGTKDLPTVQ 3196

Qy   1342 RGKALGSLGG--SPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETG-ESP 1398
Db   3197 TGD-IPPLSGVKQISCPDSSEPA---VQVQLDFSTLTRS-----------VYSDRGDDSP 3241

Qy   1399 ECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGP-------ESQAESIPI----IVTPAPE 1447
Db   3242 DS-----SPEEQKSVI-----------EIPTAPMENVPFTESKSK-IPVRTMPTSTPAPP 3284

Qy   1448 STLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL-KAEPEILELESKSNKI 1506
Db   3285 SA---EYESSVSEDFLSSVDEENKAD---EAKPKSKLPVKVPLQRVEQQLSDLDTSVQKT 3338

Qy   1507 VL----NVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPV 1562
Db   3339 VAPQGQDMASIAPDNRSKSES--DASSLDSKTKCPV-------KTRSYTETETESRER-- 3387

Qy   1563 PQPREDLQVLTVLEAWAQPRKCLPRLQLKA 1592
Db   3388 ---AEELE-LESEEGATRPKILTSRLPVKS 3413
```

```
RESULT 4
MAPB_HUMAN
ID   MAPB_HUMAN     STANDARD;      PRT;  2468 AA.
AC   P46821;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 1B (MAP 1B) [Contains: MAP1 light chain
DE   LC1].
GN   MAP1B.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Fetal brain;
RX   MEDLINE=95104835; PubMed=7806212;
RA   Lien L.L., Feener C., Fischbach N., Kunkel L.M.;
RT   "Cloning of human microtubule-associated protein 1B and the
RT   identification of a related gene on chromosome 15.";
RL   Genomics 22:273-280(1994).
CC   -!- FUNCTION: THE FUNCTION OF BRAIN MAPS IS ESSENTIALLY UNKNOWN.
CC       PHOSPHORYLATED MAP1B MAY PLAY A ROLE IN THE CYTOSKELETAL CHANGES
CC       THAT ACCOMPANY NEURITE EXTENSION. POSSIBLY MAP1B BINDS TO AT LEAST
CC       TWO TUBULIN SUBUNITS IN THE POLYMER, AND THIS BRIDGING OF SUBUNITS
CC       MIGHT BE INVOLVED IN NUCLEATING MICROTUBULE POLYMERIZATION AND IN
CC       STABILIZING MICROTUBULES.
CC   -!- SUBUNIT: 3 DIFFERENT LIGHT CHAINS, LC1, LC2 AND LC3, CAN ASSOCIATE
CC       WITH MAP1A AND MAP1B PROTEINS.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC       from MAP1B by proteolytic processing. It is free to associate with
CC       both MAP1A and MAP1B. It interacts with the amino-terminal region
CC       of MAP1B (By similarity).
CC   -!- SIMILARITY: TO MAP1A.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L06237; AAA18904.1; -.
DR   Genew; HGNC:6836; MAP1B.
DR   MIM; 157129; -.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 6.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN         ?   2468       MAP1 LIGHT CHAIN LC1.
FT   REPEAT     1878   1894       MAP1B 1.
FT   REPEAT     1895   1911       MAP1B 2.
FT   REPEAT     1912   1928       MAP1B 3.
FT   REPEAT     1929   1945       MAP1B 4.
FT   REPEAT     1946   1962       MAP1B 5.
FT   REPEAT     1963   1979       MAP1B 6.
FT   REPEAT     1997   2013       MAP1B 7.
FT   REPEAT     2014   2030       MAP1B 8.
FT   REPEAT     2031   2047       MAP1B 9.
FT   REPEAT     2048   2064       MAP1B 10.
FT   DOMAIN      589    790       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                KKEE AND KKEI/V REPEATS).
SQ   SEQUENCE   2468 AA;  270618 MW;  540839CBDF09D461 CRC64;

  Query Match             6.1%;  Score 492;  DB 1;  Length 2468;
  Best Local Similarity   20.3%;  Pred. No. 4.9e-10;
  Matches  368;  Conservative  266;  Mismatches  694;  Indels  482;  Gaps   75;

Qy     45 ADPATKL-----PQKNGQLSSVNGVAEQGDVH------VQEENQEGQEEEVVDEDVGQRE 93
Db    478 ANPAEKIIRVLFPGNSTQYNILEGLEKLKHLDFLKQPLATQKDLTGQVPTPVVKQTKLKQ 537

Qy     94 SEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVG 153
Db    538 RADSRES---LKPAAKPLPSKSVRKESKEETPEV-----TKVNHVE---KPPKVESKE-- 584

Qy    154 FKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPS-VETAVGES 212
Db    585 ----------KVMVKKDKPVKTETKPSVTEKE-----------VPSKEEPSPVKAEVAE- 622

Qy    213 ASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESP 272
Db    623 -------KQATDVKPKAAKEKTVKKETKVKPEDKKEEKEKPKKEVAKKEDKTPIKKEEKP 675
```

```
Qy     273 SSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTE 332
Db     676 K---KEEVKKEVKKEI-------KKEEKKEPKKEVKKETPPKEVKKEVKK-----EEKKE 720

Qy     333 PASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQE 392
Db     721 VKKEEKEPKKEI--KKLPKDAKKSSTPL---------SEAKKPAALKPKVPKKEESVKKD 769

Qy     393 VVAEVHVSTVEKTEEEQGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db     770 SVAAGKPKEKGKIKVIKKEGKAAEAVAAAVGTGATTAAVMAAAGIAA---IGPAKELEAE 826

Qy     453 REMCVSGGDHTQ-LTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGL 511
Db     827 RSLMSSPEDLTKDFEELKAEEVDVTK---DIKPQLELIEDEEKLK-ETEPVEAYVIQKER 882

Qy     512 KKLSGKKQKGKRG-------GGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETT 564
Db     883 EVTKGPAESPDEGITTTEGEGECEQTPEELEPVEKQGVD--DIEKFEDEGAGFEESSET- 939

Qy     565 CLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELE- 623
Db     940 ----GDYEEKAETEEAEEPEEDGEEH------------VCVSASKHSPTEDEESAKAEA 982

Qy     624 ----KVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTS----------VSWEAL 669
Db     983 DAYIREKRESVASGDDRAEEDMDEAIEKGEAEQSEEEADEEDKAEDAREEEYEPEKMEAE 1042

Qy     670 ICVGSSKKRARKASSSDDEGG-----PRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQ 724
Db    1043 DYVMAVVDKAAEAGGAEEQYGFLTTPTKQLGAQSPGREPASSIHD---ETLPGGSESEAT 1099

Qy     725 AQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLST------- 777
Db    1100 ASDEENRE--DQPEEFTATSGY-----------TQSTIEISSEPTPMDEMSTPRDVMSD 1145

Qy     778 ----EIEPSRREESWVSIKK--------------------FIPGRRKKRADGKQEQATVE 812
Db    1146 ETNNEETESPSQEFVNITKYESSLYSQEYSKPADVTPLNGFSEGSKTDATDGKDYNASAS 1205

Qy     813 DSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTV 872
Db    1206 TISPP------------------SSMEEDKFSRSA-------LRDAYCSEVKAST----- 1235

Qy     873 SVAVIDGTRAVTSVE---ERSPSWISASVTEPLEHT-AGE-----AMPPVE-EVTEKDII 922
Db    1236 TLDIKDSISAVSSEKVSPSKSPS-LSPSPPSPLEKTPLGERSVNFSLTPNEIKVSAEAEV 1294

Qy     923 AEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTAT------------ETSEALRTEEVT 970
Db    1295 APVSPEVTQEVVEEHCASPEDKTLEVVSPSQSVTGSAGHTPYYQSPTDEKSSHLPTEVI- 1353

Qy     971 EASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQA---VAD 1027
Db    1354 ------EKPPAVPVSFEFSDAKDENERAS-VSPMDEPVPDSESPIEKVLSPLRSPPLIGS 1406

Qy    1028 KVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQG 1087
Db    1407 ESAYESFLSADDKASGRGAES-----PFEEKSGKQGSPDQ-----VSPVSEMTSTSLYQD 1456

Qy    1088 SETGQAT-----PESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPE------------- 1129
Db    1457 KQEGKSTDFAPIKEDFGQEKKTDDVEAMSSQPALALDERKLGDVSPTQIDVSQFGSFKED 1516

Qy    1130 -----SSETLTD---------------SETNGSTPLADSDTADGTQQDETIDS------- 1162
Db    1517 TKMSISEGTVSDKSATPVDEGVAEDTYSHMEGVASVSTASVATSSFPEPTTDDVSPSLHA 1576

Qy    1163 ----------QDSKATAAVRQSQVTEEEAATAQKEE-------------PSTLPNNVPAQ 1199
Db    1577 EVGSPHSTEVDDSLSVSVVQTPTTFQETEMSPSKEECPRPMSISPPDFSPKTAKSRTPVQ 1636

Qy    1200 EEHGEEPGRDV---LEPTQQELT-----------AAAVPVLAKTEVGQEGEVDWLDGEKV 1245
Db    1637 DHRSEQSSMSIEFGQESPEQSLAMDFSRQSPDHPTVGAGVLHITENGPT-EVDYSPSDM- 1694

Qy    1246 KEEQEVFVHSGPNSQKAADVTYD---SEVMGVAGCQEKEST------------------- 1283
Db    1695 --QDSSLSHKIPPMEEPS-YTQDNDLSELISVSQVEASPSTSSAHTPSQIASPLQEDTLS 1751

Qy    1284 -----------------EVQSLSLEEGEMETDVE---------------------KEKR 1304
Db    1752 DVAPPRDMSLYASLTSEKVQSLEGEKLSPKSDISPLTPRESSPLYSPTFSDSTSAVKEKT 1811

Qy    1305 ETKPEQVSEEGEQETAAP------------EHEGTYGKPVLTLDMPSSERGKALGSLGGS 1352
Db    1812 ATCHSSSSPPIDAASAEPYGFRASVLFDTMQHHLALNRDLSTPGLEKDSGGKTPGDFSYA 1871

Qy    1353 PSLPDQDKAGCIE--VQVQSLDTTVTQTAEA----VEKVIETVVI-SETGESPECVGAHL 1405
Db    1872 YQKPEETTRSPDEEDYDYESYEKT-TRTSDVGGYYYEKIERTTKSPSDSGYSYETIG--- 1927

Qy    1406 LPAEKSSAT--GGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQR 1463
Db    1928 ----KTTKTPEDGDYSYEIIEKTTRT-PEEGGYSYDI----SEKTTSPPEVSG-YSYEKT 1977

Qy    1464 ERSEE---------EDKPDAGPDADGKESTAIEKVLK----AEPEILELESKSNKIVLNV 1510
Db    1978 ERSRRLLDDISNGYDDSEDGGHTL-GDPSYSYETTEKITSFPESEGYSYETST------- 2029

Qy    1511 IQTAVDQFARTETAPETHAY-----DSQTQVPACRLDSRE-PNRCWTKMKDAKMKHPVPQ 1564
Db    2030 ---------KTTRTPDTSTYCYETAEKITRTPQASTYSYETSDLCYT----AEKKSP-SE 2075

Qy    1565 PREDLQVLTV 1574
Db    2076 ARQDVDLCLV 2085
```

RESULT 5
MAPB_RAT
ID MAPB_RAT STANDARD; PRT; 2459 AA.
AC P15205; Q62958; Q9ER21; Q9QW92;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Microtubule-associated protein 1B (MAP 1B) (Neuraxin) [Contains: MAP1
DE light chain LC1].
GN MAP1B.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE OF 1-142 FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Testis;
RX MEDLINE=96257242; PubMed=8666295;
RA Liu D., Fischer I.;
RT "Isolation and sequencing of the 5' end of the rat microtubule-
RT associated protein (MAP1B)-encoding cDNA.";
RL Gene 172:307-308(1996).
RN [2]
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zauner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
RT recombinant rat MAP 1B.";
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Spinal cord;
RX MEDLINE=90059871; PubMed=2555150;
RA Rienitz A., Grenningloh G., Hermans-Borgmeyer I., Kirsch J.,
RA Littauer U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;

201

```
RT   "Neuraxin, a novel putative structural protein of the rat central
RT   nervous system that is immunologically related to microtubule-
RT   associated protein 5.";
RL   EMBO J. 8:2879-2888(1989).
RN   [4]
RP   DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX   MEDLINE=97405699; PubMed=9260743;
RA   Ma D., Nothias F., Boyne L.J., Fischer I.;
RT   "Differential regulation of microtubule-associated protein 1B (MAP1B)
RT   in rat CNS and PNS during development.";
RL   J. Neurosci. Res. 49:319-332(1997).
CC   -!- FUNCTION: The function of brain MAPS is essentially unknown.
CC       Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC       that accompany neurite extension. Possibly MAP1B Binds to at least
CC       two tubulin subunits in the polymer, and this bridging of subunits
CC       might be involved in nucleating microtubule polymerization and in
CC       stabilizing microtubules.
CC   -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC       with MAP1A and MAP1B proteins.
CC   -!- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC       cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC       heart or muscle.
CC   -!- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC       nerve levels are high early in development but decrease during
CC       postnatal development and are low in adults. In dorsal root
CC       ganglia levels remain high throughout development.
CC   -!- INDUCTION: By nerve growth factor.
CC   -!- DOMAIN: Has a highly basic region with many copies of the sequence
CC       KKEE and KKEI/V, repeated but not at fixed intervals, which is
CC       responsible for the binding of MAP1B to microtubules.
CC   -!- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC       from MAP1B by proteolytic processing. It is free to associate with
CC       both MAP1A and MAP1B. It interacts with the amino-terminal region
CC       of MAP1B (By similarity).
CC   -!- PTM: Phosphorylated.
CC   -!- SIMILARITY: TO MAP1A.
CC   -!- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC       2459) was originally described as neuraxin in Ref.3.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U52950; AAB17068.1; -.
DR   EMBL; X60370; CAC16162.1; -.
DR   EMBL; X16623; CAA34620.1; ALT_SEQ.
DR   PIR; S06017; S06017.
DR   InterPro; IPR000102; MAP1B_neuraxin.
DR   Pfam; PF00414; MAP1B_neuraxin; 10.
DR   PROSITE; PS00230; MAP1B_NEURAXIN; 8.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN          ?    2459       MAP1 LIGHT CHAIN LC1.
FT   REPEAT      1869    1885       MAP1B 1.
FT   REPEAT      1886    1902       MAP1B 2.
FT   REPEAT      1903    1919       MAP1B 3.
FT   REPEAT      1920    1936       MAP1B 4.
FT   REPEAT      1937    1953       MAP1B 5.
FT   REPEAT      1954    1970       MAP1B 6.
FT   REPEAT      1988    2004       MAP1B 7.
FT   REPEAT      2005    2021       MAP1B 8.
FT   REPEAT      2022    2038       MAP1B 9.
FT   REPEAT      2039    2055       MAP1B 10.
FT   DOMAIN       559    1035       GLU-RICH.
FT   DOMAIN       588     786       LYS-RICH (HIGHLY BASIC, CONTAINS MANY
FT                                  KKEE AND KKEI/V REPEATS).
FT   DOMAIN      2224    2312       LYS-RICH.
FT   CONFLICT     127     127       M -> V (IN REF. 1).
FT   CONFLICT     140     140       T -> S (IN REF. 1).
FT   CONFLICT    2112    2112       R -> K (IN REF. 3).
FT   CONFLICT    2169    2169       L -> I (IN REF. 3).
SQ   SEQUENCE   2459 AA;  269497 MW;  2E3F6872DEDB8BA2 CRC64;

  Query Match             5.8%;  Score 470;  DB 1;  Length 2459;
  Best Local Similarity   20.2%;  Pred. No. 2.8e-09;
  Matches  328;  Conservative  234;  Mismatches  602;  Indels  458;  Gaps   62;

Qy         45 ADPATKL-----PQKNGQLSSVNGVAEQGDVH------VQEENQEGQEEEVVDEDVGQRE 93
              |:|| |:      |  : | : : |: :   :         :::  ||         :  |  ::
Db        477 ANPAEKIIRVLFPGNSTQYNILEGLEKLKHLDFLKQPLATQKDLTGQVSTPPVKQVKLKQ 536

Qy         94 SEDVRE--KDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQAND 151
                || |     : :::  |       : |: :||  |       |:: :   |      ||:
Db        537 RADSRESLKPATKPLSSKS-----VRKESKEEAPE------ATKASQVEKTPKVESKE-- 583

Qy        152 VGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGE 211
                           | ||||  | ::    :|           |  | : : | | |: | |
Db        584 -----------KVIVKKDKPGKVESKPSVT---------EKEVPSKEEQSP-VKAEVAE 621

Qy        212 SASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPES 271
               |: ||: | : :|    :|:|          :: : ::  ||: | |   :|::| | |  |
Db        622 KAATESKPKVTKDK---VVKKE-------IKTKPEEKKEEKPKKEVAKKEDKTPLKKDEK 671

Qy        272 PSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEK--E 329
              |      |     ||                     || :: :|:  | :||   :   :| |  |
Db        672 PK---KEEAKKEIKK------------EIKKEEKKELK-KEVKKETPLKDAKKEVKKDE 714

Qy        330 KTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEA 389
              | |     ||:||  ::       ::| | :|     ||  |          :| | |  :| |
Db        715 KKEVKKEEKEPKKEI--KKISKDIKK-STPLSD---------TKKPAALKPKVAKKEEPT 762

Qy        390 HQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEEL 449
               :| :|     :       | :   :     |       ||          :|        : ||:||
Db        763 KKEPIAAGKLKDKGKVKVIKKEGKTTEAAATAVGTAAVAAAAGVA-------ASGPAKEL 815

Qy        450 MKSREMCVSGGDHTQ-LTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSS 508
                  | :   |  | |:    :|   ||   : |      |   ::|::   :|::|
Db        816 EAERSLMSSPEDLTKDFEELKAEEIDVAK---DIKPQLELIEDEEKLK----------- 860

Qy        509 SGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEK 568
                                   : ||||   |   |                              | :  |
Db        861 ------------------ETEPGEAYVIQKE----------------------TEVSK 878

Qy        569 GPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSA 628
              |   |:| :|        ||::|| : |                   ||                |||| |
Db        879 GSAESPDEGI----TTTEGEGECE-----------QTP------------EELEPV--- 907

Qy        629 TLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDE 688
                                   |:|  :: ::   |     :|      |   :::|
Db        908 --------------------EKQGVDDIEKFEDEGAGFEESSEAGDYEEKA--------- 938

Qy        689 GGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWES 748
                              ||| : :| |         :|    ||:|     ||:| | ::   |:
Db        939 -----------ETEEAEEPEEDG----------EDNVSGSASKH---SPTEDEEIAKAEA 974

Qy        749 FKRLVTPRKKSKSKLEEKAEDSSVEQLST-EIEPSREESWVSIKKFIPGRRKKRADGKQE 807
                  :  :::| :  :::||:   |  |    | | | ||                     :  :  | |
Db        975 DVH-IKEKRESVASGDDRAEEDMDEALEKGEAEQSEEEG------------EEEEDKAE 1020

Qy        808 QATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKT 867
                |  ||   |  :   :|   |  |||  :     |   :: :     |           |                :
Db       1021 DAREEDHEPDKTEAED-YVMAVVDKAAEAGVTEDQYDFLGT---PAKQPGVQSPSREPAS 1076

Qy        868 LVHTVSVAVIDGTRAVTSVEER---SPSWISAS--VTEPLEHTAGEAMPPVEEVTEKDII 922
                :|   ::         :   |  ||        |  :|:      |:    :   |   |   |   |  |:|::
Db       1077 SIHDETLPGGSESEATASDEENREDQPEEFTATSGYTQSTIEISSEPTPMDEMSTPRDVM 1136

Qy        923 AEET----------------------------PVLT--QTLPEGK--DAHDDMVTSEVD 949
               :||                                     ||:      | :|     ||  |        |
Db       1137 TDETNNEETESPSQEFVNITKYESSLYSQEYSKPVVASFNGLSDGSKTDATDGR-----D 1191

Qy        950 FTSEAVTATETS-------------EALRTEEVTEASGAEETTDMVS-------AVSQLT 989
                : :  | |  :   |                    :|  |  ||       :|||         ||              |:
```

```
Db   1192 YNASASTISPPSSMEEDKFSKSALRDAYRPEETDVKTGAELDIKDVSDERLSPAKSPSLS 1251

Qy    990 DSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKAL 1049
          || :  | ||: |         | |                   :|  ::  ||  |     ::|:
Db   1252 PSPPSPIEKTPLGERSVNFSLTPNE--------------IKASAEGEATAVVSPGVTQAV 1297

Qy   1050 ---------EKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEV 1100
                   || ||   |: :              : |  |  :    : :|  || | | |
Db   1298 VEEHCASPEEKTLEVVSPSQSVTGSAGHTPYYQSPTDEKSSHLPTEVTENAQAVPVSFEF 1357

Qy   1101 PEVTADVDHVATC----QVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQ 1156
          |   : :  :         |    :  :|: ::|  |   |   ||:      |:  | | |  :
Db   1358 TEAKDENERSSISPMDEPVPDSESPIEKVLSPLRSPPLIGSESAYEDFLSADDKALGRRS 1417

Qy   1157 DETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQ 1216
          :    : ::  |    : ::| |:: :    | ::        :| :|:    |       |
Db   1418 ESPFEGKNGKQGFSDKESPVSDLTSDLYQDKQEEKRAGFIPIKEDFSPEKKASDAEIMSS 1477

Qy   1217 ELTAAAVPVLAKTEVGQEG---EVDWLDGEKVKEEQEVFVHSGPNSQKA-------ADVT 1266
          :    |      | : ::| :|    :||          ||: :: :   |    | |:       |: |
Db   1478 QSALA----LDERKLGGDGSPTQVDVSQFGSFKEDTKMSISEGTVSDKSATPVDEGAEDT 1533

Qy   1267 YDSEVMGVAGCQ------------------------EKESTEV-QSLSLEEGE----- 1294
           | | : |||                             ||||  |||:    :
Db   1534 Y-SHMEGVASVSTASVATSSFPEPTTDDVSPSLHAEVGSPHSTEVDDSLSVSVVQTPTTF 1592

Qy   1295 METDVEKEKRE-------TKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALG 1347
           ||::    | |         : |:    :    :    |      :|              |  |   |:
Db   1593 QETEMSPSKEECPRPMSISPPDFSPKTAKSRTPVQDHRSEQSSMSIEFGQESPEHSLAMD 1652

Qy   1348 SLGGSPSLPDQDKAGCI--------------EVQVQSLDTTVTQTAEAV----EKVIET 1388
              || |     || :                     ::|  ||    :   |  |          :  |
Db   1653 FSRQSPDHPTVG-AGMLHITENGPTEVDYSPSDIQDSSLSHKIPPTEEPSYTQDNDLSEL 1711

Qy   1389 VVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPES 1448
          : :|:   ||     ||  |:: :|                  ||  :: ::      || |   :
Db   1712 ISVSQVEASPSTSSAH-TPSQIAS---------------PLQEDTLSD-----VVPPRDM 1750

Qy   1449 TLHPDLQGE-ISASQRER----------SEEEDKPDAGP-----DADGKESTAIEKVLKA 1492
          :|:  |   | : : : |:               :   |   |     |          :   ||||  :    :
Db   1751 SLYASLASEKVQSLEGEKLSPKSDISPLTPRESSPTYSPGFSDSTSGAKESTAAYQTSSS 1810

Qy   1493 EP 1494
           |
Db   1811 PP 1812
```

```
RESULT 6
MAPA_RAT
ID   MAPA_RAT       STANDARD;      PRT;  2774 AA.
AC   P34926;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 1A (MAP 1A) [Contains: MAP1 light chain
DE   LC2].
GN   MAP1A.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=92355629; PubMed=1379599;
RA   Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT   "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RT   one messenger RNA.";
RL   J. Biol. Chem. 267:16561-16566(1992).
CC   -!- FUNCTION: Structural protein involved in the filamentous cross-
CC       bridging between microtubules and other skeletal elements.
CC   -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC       with MAP1A and MAP1B proteins.
CC   -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC   -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC       APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC       THEIR MORPHOLOGY.
CC   -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC       FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC   -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC   -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B.
CC   -!- SIMILARITY: TO MAP1B.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M83196; AAB48069.1; -.
DR   PIR; A43359; A43359.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN      ?2465   2774       MAP1 LIGHT CHAIN LC2.
FT   DOMAIN       309    496       LYS-RICH (BASIC).
FT   DOMAIN       336    541       11 X 3 AA REPEATS OF K-K-[DE].
FT   REPEAT       336    338       1.
FT   REPEAT       415    417       2.
FT   REPEAT       420    422       3.
FT   REPEAT       424    426       4.
FT   REPEAT       427    429       5.
FT   REPEAT       431    433       6.
FT   REPEAT       436    438       7.
FT   REPEAT       440    442       8.
FT   REPEAT       444    446       9.
FT   REPEAT       449    451       10.
FT   REPEAT       539    541       11.
SQ   SEQUENCE   2774 AA;  299526 MW;  3DEF74427BA9D7D7 CRC64;

  Query Match             5.4%;  Score 433.5;  DB 1;  Length 2774;
  Best Local Similarity   22.0%;  Pred. No. 5.5e-08;
  Matches  388;  Conservative  231;  Mismatches  625;  Indels  523;  Gaps   89;

Qy    103 VEEMAANSTAVEDIT-KDGQE---------ETSEIIEQIPASENNVEEMVQ---PAESQA 149
          ::: | || |   |    :|:|             : ::   :||   |  |::|:    |   :
Db    216 MQKWAGNSKAKTGIVLANGKEAEISVPYLTSITALVVWLPA--NPTEKIVRVLFPGNAPQ 273

Qy    150 NDV--GFKKV--FKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSV 205
          | :   | :|:    |: :    :||    :     |    |         :: |: |       :
Db    274 NKILEGLEKLRHLDFLRYPVATQKDLAAGAVPANLKPSKIKHRADSKESLKA------AP 327

Qy    206 ETAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEP 265
          :||| : |  :|  |::  ::     | :| : |                     ||: |||
Db    328 KTAVSKLAKREEVLEEGAKEARSELAKELAKT--------------------EKKAKEP 366

Qy    266 T-KSPESPSSP-----VNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQE 319
          : | || ||            :||   : |:          ||          ||   :: :    |::|::|
Db    367 SEKPPEKPSKSERVRGESSEALKAEKRRLIKDKAG-------KKHLKEKISKLEEKKDKE 419

Qy    320 AEKVDEEEKE-KTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPL 378
          ::: :| || | |   |::       |:                :|     :::   |:     |
Db    420 KKEIKKERKELKKEEGRKEEKKDAKKDEKR------------KDTKPEVKKLSKPDLKPF 467

Qy    379 ATEVFDE----------KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESL 428
           ||           |:: :  |  :|: :|  |  |   |         | |  |
Db    468 TPEVRKTLYKAKAPGRVKVDKGRAARGEKELSSEPRTPPAQKG---AAPPAAVSGHREL- 523

Qy    429 PPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLP----------- 476
             |: |::: |:     |||  :         |    :   |      || ||
Db    524 ---ALSSPEDLTQD---FEELKREER-----GLLAEQRDTGLGEKPLPADATEQGHPSAA 572

Qy    477 ---KHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGE 533
              | | | | | :  :: : |  ||  |                |   :|    |    |:  |
```

```
Db    573 IQVTQPSGPVLEGEHVEREKEV-VPDSPGDK-----------GSTNRGPDSGAEVEKEKE 620

Qy    534 YQHIHTESPESADEQKGESSASSPEEPE---ETTCLEKGPLEA-----PQDGEAEEGTTS 585

Db    621 ----TWEERKQREAELGPENTAAREESEAEVKEDVIEKAELEEMEETHPSDEEGEETKAE 676

Qy    586 DGEKK--REGITPWASFKKMVTPKKRVR--------RPSESDKEEELEKVKSATLSSTDS 635

Db    677 SFYQKHTQEAL--------KASPKSREALGGRDLGFQGKAPEKETASFLSSLATPAGATE 728

Qy    636 TVSEMQDEV--------KTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDD 687

Db    729 HVSYIQDETIPGYSETEQTISDEEIHDEPDER----------------PAPPRFPTSTYD 772

Qy    688 EGGPRTLGGDSHRAEEASKDKEAGTDAVPAS---TQEQDQAQGSSSPEPAGSP-SEGEGV 743

Db    773 LSGPEGPG-----PFEAS---QAADSAVPASSSKTYGAPETELTYPPNMVAAPLAEEEHV 824

Qy    744 STWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVS-----IKKFIPGRR 798

Db    825 SSATS----ITECDKLSSFATSVAEDQSVASLTA---PQTEETGKSSLLLDTVTSIPSSR 877

Qy    799 KKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGA- 857

Db    878 TEATQGLDYVPSAGTISPTSSLEEDKGFKS--PPCEDFSVTGES-EKKGETVGRGLSGEK 934

Qy    858 --------VYVSEELSKTLV-------HTVSVAVIDGTRAVTSVEER--SPSWISASVTE 900

Db    935 AVGKEEKYVVTSEKLSGQYAAVFGAPGHTLP----PGEPALGEVEERCLSPDDSTVKMAS 990

Qy    901 P-----------------------------------LEHTAGEAMPPVEEVTEK------- 919

Db    991 PPPSGPPSAAHTPFHQSPVEDKSEPRDFQEDSWGETKHSPGVSKEDSEEQTVKPGPEEGT 1050

Qy    920 ------------------DIIAEETPVLTQTLPE------------GKDAHDDMVTSEV 948

Db   1051 SEEGKGPPTRSPQAQDMPVSIAGGQTGCTIQLLPEQDKAIVFETGEAGSNLGAGTLPGEV 1110

Qy    949 DFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTT-EEATPVQEVESG 1007

Db   1111 RTSTEEATEPQKDEVLR---FTDQSLSPEDAESLSVLSVV--SPDTTKQEATP-----RS 1160

Qy   1008 VLDTEEEE-----------RQTQAILQAVADKVKEESQVPATQTVQRTGSKAL--EKVEE 1054

Db   1161 PCSLKEQQPHKDLWPMVSPEDTQSL------SFSEES--PSKETSLDISSKQLSPESLGT 1212

Qy   1055 VEEDSEVLASEKEKDVMPKGPVQEAGAE--HLAQGS--ETGQAT--PESLEVP------- 1101

Db   1213 LQFGELNLGKEE------RGPVMKAEDDSCHLAPVSIPEPHRATVSPSTDETPAGTLPGG 1266

Qy   1102 ----------------EVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPL 1145

Db   1267 SFSHSALSVDRKHSPGEITGPGGHFMT-----------------SDSSLTKSPESLSSPA 1309

Qy   1146 AD------SDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKE------EPSTLP 1193

Db   1310 MEDLAVEWEGKAPGKEKEPELKSETRQ-----QKGQILPEKVAVVEQDLIIHQKDGALDE 1364

Qy   1194 NNVPA--QEEHGEEPGRDVLE-PTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQE 1250

Db   1365 ENKPGRQQDKTPEQKGRDLDEKDTAAELDKGPEP--------KEKDLD-------REDQG 1409

Qy   1251 VFVHSGPNSQK-AADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEM-ETDVEKEKRETKP 1308

Db   1410 --QRAGPPAEKDKASEQRDTDL------QQTQATEPRDRAQERRDSEEKDKSLELRDRTP 1461

Qy   1309 EQVSEEGEQETAAPEHEGTYGKPVLTLDMP--------------SSERGKAL----GSLG 1350

Db   1462 EEKDRILVQEDRAPEH--SIPEPTQTDRAPDRKGTDDKEQKEEASEEKEQVLEQKDWALG 1519

Qy   1351 GSPSLPDQDKAGCIEVQVQSL--DTTVTQTAEAVEKVI----ETVVISETGESPECV--- 1401

Db   1520 KEGETLDQE-ARTAEQKDETLKEDKTQGQKSSFVEDKTTTSKETVLDQKSAEKADSVEQQ 1578

Qy   1402 -GAHL-----LPAEKSSATGG--------HW----TLQHAEDTVPLGPE---SQAESIPI 1440

Db   1579 DGAALEKTRALGLEESPAEGSKAREQEKKYWKEQDVVQGWRETSPTRGEPVGGQKEPVPA 1638

Qy   1441 IVTPAPESTLH----PDLQGEISASQRERS---------EEEDKPDAGPD-ADGKESTAIE 1487

Db   1639 WEGKSPEQEVRYWRDRDITLQQDAYWRELSCDRKVWFPHELDGQGARPRYCEERESTFLD 1698

Qy   1488 KVLKAEPEILELESKSNKIVLNVIQTAVDQFARTETAPETHAY-DSQTQVPACRLDSREP 1546

Db   1699 EG-PDEQEITPLQ------------------HTPRSPWTSDFKDFQEPLPQKGLEVER- 1737

Qy   1547 NRCWTKMKDAKMKHPVPQPREDLQVLT 1573

Db   1738 ---WL------AESPVGLPPEEEDKLT 1755
```

```
RESULT 7
MAPA_HUMAN
ID   MAPA_HUMAN     STANDARD;      PRT;  2805 AA.
AC   P78559; Q15882; Q12973; Q9UJT4; O95643;
DT   15-JUL-1998 (Rel. 36, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 1A (MAP 1A) (Proliferation-related
DE   protein p80) [Contains: MAP1 light chain LC2].
GN   MAP1A.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=97001161; PubMed=8812494;
RA   Fink J.K., Jones S.M., Esposito C., Wilkowski J.;
RT   "Human microtubule-associated protein 1a (MAP1A) gene: genomic
RT   organization, cDNA sequence, and developmental- and tissue-specific
RT   expression.";
RL   Genomics 35:577-585(1996).
RN   [2]
RP   SEQUENCE OF 1-1825 FROM N.A.
RC   TISSUE=Ovarian carcinoma;
RA   Chen Z.C., Fadiel A., Naftolin F.;
RT   "Identification of a novel protein (P80) in ovarian carcinoma cells.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 78-1687 FROM N.A.
RC   TISSUE=Brain;
RA   Ohtani K., Rutherford T., Sakamoto H., Naftolin F.;
RT   "Microtubule associated protein 1A (MAP1A) in human brain - DNA
RT   sequence and physiological role.";
RL   Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 134-419 FROM N.A.
RC   TISSUE=Fetal muscle;
RA   Chiannilkulchai N., Pasturaud P., Richard I., Auffray C.,
RA   Beckmann J.S.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE OF 1607-1883 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=95356255; PubMed=7629894;
RA   Fukuyama R., Rapoport S.I.;
RT   "Brain-specific expression of human microtubule-associated protein 1A
RT   (MAP1A) gene and its assignment to human chromosome 15.";
RL   J. Neurosci. Res. 40:820-825(1995).
CC   -!- FUNCTION: Structural protein involved in the filamentous cross-
CC       bridging between microtubules and other skeletal elements.
CC   -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC       with MAP1A and MAP1B proteins.
CC   -!- TISSUE SPECIFICITY: BRAIN.
CC   -!- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
```

```
CC       FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC   -!- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC   -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC       FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC       BOTH MAP1A AND MAP1B.
CC   -!- SIMILARITY: TO MAP1B.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U38291; AAB41132.1; -.
DR   EMBL; U38292; AAB41133.1; -.
DR   EMBL; AF200415; AAF08305.2; -.
DR   EMBL; U80458; AAD00355.1; -.
DR   EMBL; Z47038; CAA87104.1; -.
DR   EMBL; U14577; AAA81362.1; -.
DR   Genew; HGNC:6835; MAP1A.
DR   MIM; 600178; -.
KW   Microtubules; Repeat; Phosphorylation.
FT   CHAIN      ?2490   2805       MAP1 LIGHT CHAIN LC2.
FT   DOMAIN       309    496       LYS-RICH (BASIC).
FT   DOMAIN       415    541       9 X 3 AA REPEATS OF K-K-[DE].
FT   REPEAT       415    417       1.
FT   REPEAT       420    422       2.
FT   REPEAT       427    429       3.
FT   REPEAT       431    433       4.
FT   REPEAT       436    438       5.
FT   REPEAT       440    442       6.
FT   REPEAT       444    446       7.
FT   REPEAT       449    451       8.
FT   REPEAT       539    541       9.
FT   CONFLICT      72     72       F -> L (IN REF. 1).
FT   CONFLICT     134    135       VV -> IP (IN REF. 4).
FT   CONFLICT     249    249       A -> G (IN REF. 4).
FT   CONFLICT     263    263       V -> A (IN REF. 4).
FT   CONFLICT     296    296       Q -> H (IN REF. 3).
FT   CONFLICT     311    311       S -> G (IN REF. 4).
FT   CONFLICT     324    324       K -> Q (IN REF. 3).
FT   CONFLICT     335    336       AK -> ST (IN REF. 1).
FT   CONFLICT     353    353       A -> S (IN REF. 1).
FT   CONFLICT     357    357       A -> S (IN REF. 1).
FT   CONFLICT     364    364       K -> Q (IN REF. 1).
FT   CONFLICT     414    419       EKKDKE -> KKKRNS (IN REF. 4).
FT   CONFLICT     424    424       K -> P (IN REF. 1).
FT   CONFLICT     426    426       E -> D (IN REF. 3).
FT   CONFLICT     431    431       K -> Q (IN REF. 3).
FT   CONFLICT     439    439       E -> D (IN REF. 3).
FT   CONFLICT     444    444       K -> R (IN REF. 3).
FT   CONFLICT     452    453       TK -> SS (IN REF. 3).
FT   CONFLICT     457    457       K -> R (IN REF. 3).
FT   CONFLICT     682    682       Q -> P (IN REF. 1).
FT   CONFLICT    1025   1025       Q -> K (IN REF. 1).
FT   CONFLICT    1303   1313       KYLPGAITSPD -> EVLTWGDHQALN (IN REF. 3).
FT   CONFLICT    1335   1341       MISSING (IN REF. 3).
FT   CONFLICT    1368   1368       Q -> T (IN REF. 3).
FT   CONFLICT    1470   1470       A -> T (IN REF. 3).
FT   CONFLICT    1650   1650       W -> C (IN REF. 1).
FT   CONFLICT    1690   1690       A -> S (IN REF. 1).
FT   CONFLICT    1714   1714       G -> V (IN REF. 1).
FT   CONFLICT    1869   1869       E -> A (IN REF. 5).
FT   CONFLICT    1879   1883       GTPEY -> AHSRV (IN REF. 5).
SQ   SEQUENCE   2805 AA;  306392 MW;  F3ED0A3165993B2E CRC64;

  Query Match             5.2%;  Score 421;  DB 1;  Length 2805;
  Best Local Similarity   20.8%;  Pred. No. 1.5e-07;
  Matches  349;  Conservative  237;  Mismatches  671;  Indels  420;  Gaps   72;

Qy     103 VEEMAANSTAVEDIT-KDGQE---------ETSEIIEQIPASENNVEEMVQ---PAESQA 149

Db     216 MQKWAGNSKAKTGIVLPNGKEAEISVPYLTSITALVVWLPA--NPTEKIVRVLFPGNAPQ 273

Qy     150 NDV--GFKKV--FKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSV 205

Db     274 NKILEGLEKLRHLDFLRYPVATQKDLASGAVPTNLKPSKIKQRADSKESLKA------TT 327

Qy     206 ETAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEP 265

Db     328 KTAVSKLAKREEVVEEGAKEARSELAKELAKT------------EKKAKESSE------- 368

Qy     266 TKSPESPSSP--VNSETTSSFKKFFTHGWAGWRKKTSFKKSK---EDDLETAEKRKEQEA 320

Db     369 -KPPEKPAKPERVKTESSEALK-------AEKRKLIKDKVGKKHLKEKISKLEEKKDKEK 420

Qy     321 EKVDEEEKE--KTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPL 378

Db     421 KEIKKERKELKKDEGRKEEKKDA-------------KKEEKRKDTKPELKKISKPDLKPF 467

Qy     379 ATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQE 438

Db     468 TPEV--------RKTLYKAKVPGRVKIDRSRAIRGE---------------KELSSEPQT 504

Qy     439 VPQE--AEPAEELMKSREMCVSG-GDHTQ-LTDLSPEEKTL--PKHPEGIVSEVEMLSSQ 492

Db     505 PPAQKGTVPLPTISGHRELVLSSPEDLTQDFEEMKREERALLAEQRDTGLGDKPFPLDTA 564

Qy     493 ER----IKVQGSP------------LKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQH 536

Db     565 EEGPPSTAIQGTPPSVPGLGQEEHVMKEKELVPEVPEEQGSKDRGLDSGAETEE--EKDT 622

Qy     537 IHTESPESADEQKGESSASSPEEPE-ETTCLEKGPLEA-----PQDGEAEEGTTSDGEKK 590

Db     623 WEEKKQREAERLPDRTEAREESEPEVKEDVIEKAELEEMEEVHPSDEEEEDATKAEGFYQ 682

Qy     591 REGITPWASFKKMVTPKKRVRRPSESDKEEEL-------EKVKSATLSSTDST------V 637

Db     683 ----------KHMQEPLKVTPRSREAFGGRELGLQGKAPEKETSLFLSSLTTPAGATEHV 732

Qy     638 SEMQDEV--------KTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEG 689

Db     733 SYIQDETIPGYSETEQTISDEEIHDEPEER----------------PAPPRFHTSTYDLP 776

Qy     690 GPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEP---------AGSPSEG 740

Db     777 GPEGAG-----PFEASQPADS---AVPATS-----GKVXGTPETELTYPTNIVAAPLAEE 823

Qy     741 EGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVS-----IKKFIP 795

Db     824 EHVSSATS----ITECDKLSSFATSVAEDQSVASLTA---PQTEETGKSSLLLDTVTSIP 876

Qy     796 GRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAV-EREKMEAQGNTELPQL 854

Db     877 SSRTEATQGLDYVPSAGTISPTSSLEEDKGFKS-PPCEDFSVTGESEKR----------- 924

Qy     855 LGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPS-----WISASVTEPLEHTAGEA 909

Db     925 ------GEIIGKGL---------SGERAVEEEEEETANVEMSEKLCSQYGTPVFSAPGHA 969

Qy     910 MPPVE----EVTEKDIIAEETPV-LTQTLPEGKDA--HDDMVTSEVDFTSEAVTATET-- 960

Db     970 LHPGEPALGEAEERCLSPDDSTVKMASPPPSGPPSATHTPFHQSPVEEKSEPQDFQEADS 1029

Qy     961 -SEALRTEEVTEASGAEETTDMVSAVSQLTD---------SPDTTEEATPVQEVESG---- 1007

Db    1030 WGDTKRTPGVGKEDAAEETVKPGPEEGTLEKEEKVPPPRSPQAQEAPVNIDEGLTGCTIQ 1089

Qy    1008 ---------VLDTEEEERQTQAILQAVA-----DKVKEESQVPATQTVQRTGSKALEKVE 1053

Db    1090 LLPAQDKAIVFEIMEAGEPTGPILGAEALPGGLRTLPQEPGKPQKDEVLRYPDRSLS--- 1146

Qy    1054 EVEEDSEVLA-------SEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTAD 1106
```

```
Db   1147 --PEDAESLSVLSVPSPDTANQEPTPKSPC--GLTEQYLHKDRWPEVSPEDTQSLSLSEE 1202

Qy   1107 VDHVATCQVIKLQQLMEQAVAPESSETLTDSETN-------------------------- 1140
               |    :   :||      :||  ||     |  |
Db   1203 SPSKETSLDVSSKQL-----SPESLGTLQFGELNLGKEEMGHLMQAEDTSHHTAPMSVPE 1257

Qy   1141 ----GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQV--------------TEEEAA 1182
              ::|  |   |      |  |  | || |  |: |:                     :|
Db   1258 PHAATASPPTDGTTRYSAQTDITDDSLDRKSPASSFSHSTPSGNGKYLPGAITSPDEHIL 1317

Qy   1183 TAQ---KEEPSTLP------------NNVP------AQEEHGEEPGRDVLEPTQQELTAA 1221
          |       :  | :||               : ||       ::::     || :||:   :  |
Db   1318 TPDSSFSKSPESLPGPALEDIAIKWEDKVPGLKDRTSEQKKEPEPKDEVLQQKDKTLEHK 1377

Qy   1222 AVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKE 1281
             |    | :  |: |     :   | ||:: :        |          ||    |   |: :
Db   1378 EVVEPKDTAIYQKDEALHVKNEAVKQQDKALEQKG----------RDLEQKDTALEQKDK 1427

Qy   1282 STEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSE 1341
          : |  :     ||     | |    |:::   ||:   :   ||:   | |  :    |   :        :
Db   1428 ALEPKDKDLE----EKDKALEQKDKIPEEKDKALEQKDTALEQK---DKALEPKDKDLEQ 1480

Qy   1342 RGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQ--TAEAVEKVIETVVISETGESPE 1399
          : :  |         |      ::|||   :: :|:|::    :    ||    :: :|       :| ::||
Db   1481 KDRVLEQKEKIPE--EKDKA--LDQKVRSVEHKAPEDTVAEMKDRDLE-----QTDKAPE 1531

Qy   1400 CVGAHLLPAEKSSAT--------GGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLH 1451
                 |       :|    :              :|  |      :: :      :: |:        |      || :
Db   1532 --QKHQAQEQKDKVSEKKDQALEQKYWALGQKDEALEQNIQALEENHQ---TQEQESLVQ 1586

Qy   1452 PDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKV--------LKAEPEILELESKS 1503
             |          ::: :   ||  |:       :  |    :||         |    |     |  |  |
Db   1587 ED------KTRKPKMLEEKSPEKVKAMEEKLEALLEKTKALGLEESLVQEGRAREQEEKY 1640

Qy   1504 NKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSRE-PNRCWTKMKDAKMK 1559
             :          |   | ::   |        |   |  : :     ||        |  ||  |     :|    ::
Db   1641 WR-----GQDVVQEWQETSPTREEPAGEQKELAPAWEDTSPEQDNRYWRGREDVALE 1692


RESULT 8
PGCV_CHICK
ID   PGCV_CHICK     STANDARD;      PRT;  3562 AA.
AC   Q90953; Q90945;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Versican core protein precursor (Large fibroblast proteoglycan)
DE   (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN   CPSG2.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=White leghorn; TISSUE=Limb bud;
RX   MEDLINE=93300846; PubMed=8314802;
RA   Shinomura T., Nishida Y., Ito K., Kimata K.;
RT   "cDNA cloning of PG-M, a large chondroitin sulfate proteoglycan
RT   expressed during chondrogenesis in chick limb buds. Alternative
RT   spliced multiforms of PG-M and their relationships to versican.";
RL   J. Biol. Chem. 268:14461-14469(1993).
CC   -!- FUNCTION: May play a role in intercellular signaling and in
CC       connecting cells with the extracellular matrix. May take part in
CC       the regulation of cell motility, growth and differentiation. Binds
CC       hyaluronic acid.
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC   -!- ALTERNATIVE PRODUCTS: At least 2 isoforms; V0 (shown here) and
CC       V1; are produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Prechondrogenic condensation area of
CC       developing limb buds.
CC   -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development
CC       (By similarity).
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X60226; CAA42787.1; -.
DR   EMBL; D13542; BAA02742.1; -.
DR   HSSP; P00740; 1EDM.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 2.
DR   ProDom; PD000918; Link; 2.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS01241; LINK; 2.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW   Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW   Hyaluronic acid; Alternative splicing.
FT   SIGNAL        1     26       POTENTIAL.
FT   CHAIN        27   3562       VERSICAN CORE PROTEIN.
FT   DOMAIN       37    136       IG-LIKE V-TYPE DOMAIN.
FT   DOMAIN      166    243       LINK 1.
FT   DOMAIN      264    345       LINK 2.
FT   DOMAIN     3254   3290       EGF-LIKE 1.
FT   DOMAIN     3292   3328       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     3341   3455       C-TYPE LECTIN.
FT   DOMAIN     3460   3518       SUSHI.
FT   DISULFID     44    129       BY SIMILARITY.
FT   DISULFID    171    242       BY SIMILARITY.
FT   DISULFID    195    216       BY SIMILARITY.
FT   DISULFID    269    344       BY SIMILARITY.
FT   DISULFID    293    314       BY SIMILARITY.
FT   DISULFID   3258   3269       BY SIMILARITY.
FT   DISULFID   3263   3278       BY SIMILARITY.
FT   DISULFID   3280   3289       BY SIMILARITY.
FT   DISULFID   3296   3307       BY SIMILARITY.
FT   DISULFID   3301   3316       BY SIMILARITY.
FT   DISULFID   3318   3327       BY SIMILARITY.
FT   DISULFID   3334   3345       BY SIMILARITY.
FT   DISULFID   3362   3454       BY SIMILARITY.
FT   DISULFID   3430   3446       BY SIMILARITY.
FT   DISULFID   3461   3504       BY SIMILARITY.
```

```
FT   DISULFID   3490   3517       BY SIMILARITY.
FT   CARBOHYD    163    163       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    235    235       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    329    329       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    529    529       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    709    709       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    948    948       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1409   1409       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1479   1479       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1523   1523       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1530   1530       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1625   1625       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1751   1751       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1988   1988       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2088   2088       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2089   2089       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2507   2507       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2642   2642       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2679   2679       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2748   2748       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2762   2762       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3069   3069       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3194   3194       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3232   3232       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   3545   3545       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    485   1411       MISSING (IN ISOFORM V1).
SQ   SEQUENCE   3562 AA;  388078 MW;  9BC566E88C1602D2 CRC64;

  Query Match              5.1%;  Score 411.5;  DB 1;  Length 3562;
  Best Local Similarity   19.7%;  Pred. No. 4e-07;
  Matches  397;  Conservative  269;  Mismatches  698;  Indels  649;  Gaps   94;

Qy       49 TKLPQ--KNGQLSSVNGVAE-QGDVHVQ---------EENQEGQEEEVVDEDVGQRESED 96
Db      433 TLLPQTVTDGEISPYDTLGRTEYDVSPRLTESTSAALEVEHTYSEAELSEE---QGRSES 489

Qy       97 VREKDRVEEMAANSTAV--------EDITK--------DGQEETSEIIEQIPASENNVEE 140
Db      490 TEDAFLTSVVFQDSTAVAKSSTGSWEDIETGDTQKHDGDNQTEQIEVGPVMTATDSLVPA 549

Qy      141 MVQPAESQANDVGFKKVFKFVGFKFT---VKKDKNEKSD----TV----QLLTVKKD--- 186
Db      550 SQRELPRTGSSVSLTKENLYLGSHSTKEPTKKSMEAKSDKKLTTVVIPKALFTDQYDLTT 609

Qy      187 EGEGAE-------------ASVGAGDHQEPSV-ETAVGESASKESELKQSTEKQEGT--L 230
Db      610 GGEGRESMYTVMPDRVSGVALVSIPESDVPAVSETLMDELAVTTG---QSSTADESTPFI 666

Qy      231 KQEQSSTEIPLQAESDQAAEEEAKD----------------------------------- 255
Db      667 KFSSSATELDNEA-SAEGSREDLKDVHLTTSSGIPVSFTLFTANETGSEVTALSESTSAP 725

Qy      256 -----------------EGE---EKQEKEPTKSPESPSSPVN-----------SETTSSFK 285
Db      726 QKFEEGITSVLHSSQQTEGSAILEKQEK--TKEPEMSTIDAKVLYITTVVPASVTAGSEG 783

Qy      286 KF----FTHG------WAGWRK-------KTSFKK----SKEDDLETAE----------- 313
Db      784 RFGSEKFTHTPPVSGMWLQTDKDQVYMTEETSHTKRIELDTEDDISGMEPTSSPGQIIEY 843

Qy      314 -KRKEQEAEKVDEEEK---EKTEPASEEQEPAEDTDQAR---------LSADYEKV---E 357
Db      844 TKHLGAPVSAVTDETKTSMETAETESDEEVVSADFDQTKGTTEVFHTSSSLDLEKFTLSK 903

Qy      358 LPLEDQVGDLEASSEEKCAPLAT------EVFDEKMEAHQEVVAEVHVSTVE-------- 403
Db      904 IPEDESSATVKSFSSSSGTVLPTAVATVLEVTDHEADETSGYVLNMTFSTPEGEQRKATE 963

Qy      404 -----KTEEEQGGGGE------AEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKS 452
Db      964 KSPATSAEDEVSTGTEISKYTMTEGGQISSVT--SAEKESVAALQE--REEQPSVGLPET 1019

Qy      453 REMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERI----------------- 495
Db     1020 KE-------PFKFTDVTEIETTVPQR-EGDTSLVPVTVGSEDIGEMQVTDHTSFDSIIHT 1071

Qy      496 --KVQGSPLKKLFSSSGLKKLSGKKQKGKRGGG-GDEEPGEYQHIHT-------ESPESA 545
Db     1072 EATVTSTKASEVFP----KELSTKDQDRELGTAMGSTLPVTSVQMHEQKTTAGFESPQTT 1127

Qy      546 DEQKGESSASSPEEPEETTCLE-----------KGPLEAPQ-----DGEAEEGTTSDGE 588
Db     1128 TQEKHDEMGSAYDEMYPATELSVPALMLTEYGQVSGPVETSTRSLHLTGTPKAETATDQE 1187

Qy      589 KK----------------------REGITPWASFKKMVTPKKRVRRPSE---------- 615
Db     1188 EKITEAVPVTFGTQAKVYESKGTTTREEDRDVGSWNSVLPPHTMLSSPSTAGSISLLTLG 1247

Qy      616 ---------SDKEEELEKVKSATLSS--TDST--VSEMQ-------DEVKTVGEEQK--- 652
Db     1248 ASPSQTPEGSGISEELEEVKTVPFSSRATDKTTVISDLTTSSISAVDKIQPTSASKPFVS 1307

Qy      653 -------PEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEAS 705
Db     1308 SKSPRIIPEE-----DEEVTSSDIIVIDESISPS-KASAEDDLTG---------KMVEPE 1352

Qy      706 KDKEAGTDA---------VPASTQEQDQAQGSSSPEPAGSPSEGEGVSTW---------- 746
Db     1353 IDKEYFTSSTATAVARPTAPPTVMEATEALQPQEVSPTSHPDSGTDIRLYVIQITGNDTD 1412

Qy      747 -------ESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTE-IEPSREESWVSIKKFIPGRR 798
Db     1413 HPVNEFLDLFSRHILPHAVDETHTD--AESAQTEPCTSDSVQDSSE--YIILDPFFP--- 1465

Qy      799 KKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQL---- 854
Db     1466 -NFMDFEEEEEDCE-------NTTDVTTP---PALQFINGKQQVTSAPKSTKAEEARSDQ 1514

Qy      855 LGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISAS--VTEPLEHTAGEAM-P 911
Db     1515 IESVAHSKNVTFSQINETNTFIISETEASGTMQPSKAGEVMGAFEVTQP---TADVAMLE 1571

Qy      912 PV----EEVTEKDIIAEETPVLTQTLPEGKDA---HDDMVTSEVDFTSEAVTATETSEAL 964
Db     1572 PVYSGESEVTTTDKYLEITSVYEQSPKKNKETVMWHG----------------------- 1608

Qy      965 RTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQA 1024
Db     1609 -----TEESSTKDTKNLLLITNE--SSGDGSTESDLSRSVFTEILTMSSHED-------- 1653

Qy     1025 VADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHL 1084
Db     1654 -SEKISHTTSVPTILSVERS---AVTAAPSADSDTATVGID-VKDLIPKG---------- 1698

Qy     1085 AQGSETGQATPESLEVPEVTADVDH------VATCQVIKLQQLMEQAVAPESSETLTDSE 1138
Db     1699 ------GTATPGNYYKSTIKLDAEFPFESNPEATSHTTKPDMTASSFIVLEGSGDVEENS 1752

Qy     1139 TNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPA 1198
Db     1753 T-----LASAMTTE-TAVAETLSVQDT----SLGSGTVLPTEISVTISEITPALPG---- 1798

Qy     1199 QEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPN 1258
Db     1799 --------GTRILYSTFDQSSEATVSTNFVSELIMEQVV----GSSVATEKKV-----ED 1841

Qy     1259 SQKAADVTYDSEVMGVAGCQEKE--------STEVQSLSLE----------EGEMETDVE 1300
Db     1842 EKEVQTTVYSSQEISTTDAKGKSELDEFGSTTNEVRTVSQEPTPLREIVPITGTMHSEIK 1901

Qy     1301 KEKRE--TKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKA-LGSLGGSPSLPD 1357
Db     1902 KVTATPFLREKLFINEGSAEEPADLFAGSPTRKVVSTDSPFTDSGSGDIDVITESATLTS 1961

Qy     1358 QDKAGCIEVQ------------------------------VQSLDTT----VTQTAEAV 1382
```

```
Db   1962 VPSRSVIETQTVKHEGNINVISVSLKNTTTEYEEHIGTGGPVTSVSSTGSDGLTEESEVA 2021

Qy   1383 EKVIETVVISETGESP--ECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
            :: | |  :|     | | |  :  |::  |  |           | |  : |  : : :
Db   2022 IEMSENVFSTENQGEPTQEAVPTYTAPSDIKSRLG-----SRREVTSHVTPVIRTKDLET 2076

Qy   1441 I-VTPAPESTLHPD------LQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAE 1493
            || :||| ::           | | |       :: |           : :: |:|  |
Db   2077 AEVTSSPESVVNNSTLDTMVTHGTIRAVAESTESKKGK-------GSFSAVSLGKILMIE 2129

Qy   1494 ---PEILELESKSNKIVLN-------------------VIQTAVDQFARTETAPE----- 1526
               | |:::| : |:: |                   :|  : |  :   :|
Db   2130 HGSGEELKVDSSTTKLMSNGPTEKLLGSHFSFFDQGSGEAETLTESFTKASVSPTGKPEP 2189

Qy   1527 --------------THAYDSQTQVPACRLDSREPNRCWTK----------MKDAKMKH- 1560
                         |||         : |||   |            : | :|:
Db   2190 QEQYGRKTVSMPSAVVHAY-----------TAEPNELVTSTEHDITSLQTVTDTEMEEK 2237

Qy   1561 -----PVPQPREDLQVLTVLEAWA-QPRKCLPR 1587
               |       :|  :   : :|   :||: ||:
Db   2238 AANELTVTSFATNLPLSEDVHSWEDRPREILPK 2270


RESULT 9
PLE1_RAT
ID   PLE1_RAT       STANDARD;      PRT;  4687 AA.
AC   P30427; O08879; O08880; O08881;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Plectin 1 (PLTN) (PCN).
GN   PLEC1.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Glial tumor;
RX   MEDLINE=91268156; PubMed=2050743;
RA   Wiche G., Becker B., Luber K., Weitzer G., Castanon M.J.,
RA   Hauptmann R., Stratowa C., Stewart M.;
RT   "Cloning and sequencing of rat plectin indicates a 466-kD polypeptide
RT   chain with a three-domain structure based on a central alpha-helical
RT   coiled coil.";
RL   J. Cell Biol. 114:83-99(1991).
RN   [2]
RP   REVISIONS.
RC   TISSUE=Glial tumor;
RX   MEDLINE=96210632; PubMed=8633055;
RA   Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT   "Human plectin: organization of the gene, sequence analysis, and
RT   chromosome localization (8q24).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN   [3]
RP   PARTIAL SEQUENCE FROM N.A. (ISOFORMS 2; 3; 4), AND TISSUE SPECIFICITY.
RC   TISSUE=Glial tumor;
RX   MEDLINE=97321050; PubMed=9177781;
RA   Elliott C.E., Becker B., Oehler S., Castanon M.J., Hauptmann R.,
RA   Wiche G.;
RT   "Plectin transcript diversity: identification and tissue distribution
RT   of variants with distinct first coding exons and rodless isoforms.";
RL   Genomics 42:115-125(1997).
CC   -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC       MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC       HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC       STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC       ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC   -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC   -!- ALTERNATIVE PRODUCTS: 4 ISOFORMS; 1 (SHOWN HERE), 2, 3 AND 4; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST EXPRESSION IN
CC       SKELETAL MUSCLE AND LOWEST IN THYMUS.
CC   -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC       VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
CC       AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC   -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC       INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC   -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 33 PLECTIN REPEATS.
CC   -!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC   -!- CAUTION: ISOFORM 4 IS A FRAGMENT AT THE N-TERMINUS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X59601; CAA42169.1; -.
DR   EMBL; U96274; AAC53209.1; -.
DR   EMBL; U96275; AAC53210.1; -.
DR   EMBL; U96276; AAC53211.1; -.
DR   PIR; A39638; A39638.
DR   PIR; S21876; S21876.
DR   HSSP; Q01082; 1BKR.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR005326; S10_plectin.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00681; Plectin; 21.
DR   Pfam; PF03501; S10_plectin; 1.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00250; PLEC; 33.
DR   SMART; SM00150; SPEC; 4.
DR   PROSITE; PS00019; ACTININ_1; FALSE_NEG.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
KW   Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW   Phosphorylation; Alternative splicing.
FT   DOMAIN        1   1473       GLOBULAR 1.
FT   DOMAIN     1474   2758       CENTRAL FIBROUS ROD DOMAIN.
FT   DOMAIN     2759   4687       GLOBULAR 2.
FT   DOMAIN      181    406       ACTIN-BINDING.
FT   DOMAIN      185    288       CH 1.
FT   DOMAIN      301    403       CH 2.
FT   REPEAT      648    722       SPECTRIN 1.
FT   REPEAT      743    827       SPECTRIN 2.
FT   REPEAT      840    933       SPECTRIN 3.
FT   REPEAT     1318   1418       SPECTRIN 4.
FT   DOMAIN     1472   1692       COILED COIL (POTENTIAL).
FT   DOMAIN     1724   2760       COILED COIL (POTENTIAL).
FT   REPEAT     2791   2828       PLECTIN 1.
FT   REPEAT     2829   2866       PLECTIN 2.
FT   REPEAT     2867   2904       PLECTIN 3.
FT   REPEAT     2905   2942       PLECTIN 4.
FT   REPEAT     2943   2980       PLECTIN 5.
FT   REPEAT     2984   3018       PLECTIN 6.
FT   REPEAT     3119   3156       PLECTIN 7.
FT   REPEAT     3157   3194       PLECTIN 8.
FT   REPEAT     3195   3232       PLECTIN 9.
FT   REPEAT     3233   3270       PLECTIN 10.
FT   REPEAT     3271   3308       PLECTIN 11.
FT   REPEAT     3311   3346       PLECTIN 12.
FT   REPEAT     3488   3525       PLECTIN 13.
FT   REPEAT     3526   3563       PLECTIN 14.
FT   REPEAT     3564   3601       PLECTIN 15.
FT   REPEAT     3602   3639       PLECTIN 16.
FT   REPEAT     3643   3677       PLECTIN 17.
FT   REPEAT     3823   3860       PLECTIN 18.
```

```
FT   REPEAT     3861   3898       PLECTIN 19.
FT   REPEAT     3899   3936       PLECTIN 20.
FT   REPEAT     3937   3974       PLECTIN 21.
FT   REPEAT     3978   4011       PLECTIN 22.
FT   REPEAT     4066   4103       PLECTIN 23.
FT   REPEAT     4104   4141       PLECTIN 24.
FT   REPEAT     4142   4179       PLECTIN 25.
FT   REPEAT     4180   4217       PLECTIN 26.
FT   REPEAT     4221   4255       PLECTIN 27.
FT   REPEAT     4268   4308       PLECTIN 28.
FT   REPEAT     4411   4448       PLECTIN 29.
FT   REPEAT     4449   4486       PLECTIN 30.
FT   REPEAT     4487   4524       PLECTIN 31.
FT   REPEAT     4525   4562       PLECTIN 32.
FT   REPEAT     4563   4600       PLECTIN 33.
FT   DOMAIN     4253   4303       BINDING TO INTERMEDIATE FILAMENTS.
FT   DOMAIN     4628   4643       4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT   MOD_RES    4542   4542       PHOSPHORYLATION (BY CDC2) (BY
FT                                SIMILARITY).
FT   VARSPLIC      1    180       MVAGMLMPLDQLRAIYEVLFREGVMVAKKDRRPRSLHPHVP
FT                                GVTNLQVMRAMTSLKARGLVRETFAWCHFYWYLTNEGIDHL
FT                                RQYLHLPPEIVPASLQRVRRPVAMVMPARRRSPHVQTMQGP
FT                                LGCPPKRGPLPAEDPAREERQVYRRKEREEGAPETPVVSAT
FT                                IVGTLARPGPEPTPAT -> MSQQRLRVPEPEGLGSKRTSS
FT                                EDNLYLAVLRASEGKK (IN ISOFORM 2).
FT   VARSPLIC      1    180       MVAGMLMPLDQLRAIYEVLFREGVMVAKKDRRPRSLHPHVP
FT                                GVTNLQVMRAMTSLKARGLVRETFAWCHFYWYLTNEGIDHL
FT                                RQYLHLPPEIVPASLQRVRRPVAMVMPARRRSPHVQTMQGP
FT                                LGCPPKRGPLPAEDPAREERQVYRRKEREEGAPETPVVSAT
FT                                IVGTLARPGPEPTPAT -> MEPSGSLFPSLVVVGHVVSLA
FT                                AVWHWRKGHRQAQDEQ (IN ISOFORM 3).
FT   VARSPLIC      1    180       MVAGMLMPLDQLRAIYEVLFREGVMVAKKDRRPRSLHPHVP
FT                                GVTNLQVMRAMTSLKARGLVRETFAWCHFYWYLTNEGIDHL
FT                                RQYLHLPPEIVPASLQRVRRPVAMVMPARRRSPHVQTMQGP
FT                                LGCPPKRGPLPAEDPAREERQVYRRKEREEGAPETPVVSAT
FT                                IVGTLARPGPEPTPAT -> DVSNGSSGSPSPGDTLPWNLG
FT                                KTQRSRRSGGGSVGNGSVLDPAERAVIRIA (IN
FT                                ISOFORM 4).
SQ   SEQUENCE   4687 AA;  533527 MW;  9966CAF71B929751 CRC64;

  Query Match             5.1%;  Score 408.5;  DB 1;  Length 4687;
  Best Local Similarity   19.8%;  Pred. No. 6.6e-07;
  Matches  317;  Conservative  264;  Mismatches  659;  Indels  361;  Gaps   64;

Qy      8 EQRSPEQPAGSD----TPSELVLSGHGPAAEASGAAGDPADADPATKLPQK------NGQ 57
Db   1487 EERLAEQQRAEERERLAEVEAALEKQRQLAEAHAQKAQAELE-ARELQRRMQEEVTRRE 1545

Qy     58 LSSVNGVAEQGDVHVQEENQ---EGQEEEVVDE----DVGQRESEDVREKDRVEEMAANS 110
Db   1546 EAAVD--AQQQKRSIQEELQHLRQSSEAEIQAKAQQVEAAERSRMRIEEEIRVVRLQLET 1603

Qy    111 TAVEDITKDGQEETSEII----EQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFT 166
Db   1604 T---ERQRGGAEDELQALRARAEEAEAQKRQAQEEAERLRRQVQDESQRKRQAEAELALR 1660

Qy    167 VKKD---KNEKSDTVQLLTVKKDEGEGAE-----ASVGAGDHQEPSVETAVG----ESAS 214
Db   1661 VKAEAEAAREKQRALQALDELKLQAEEAERWLCQAEAERARQVQVALETAQRSAEVELQS 1720

Qy    215 KESELKQSTEKQEGTLKQEQ-SSTEIPLQAE---SDQAAEEEAKDEGEEKQEKEPTKSPE 270
Db   1721 KRPSFAEKTAQLERTLQEEHVTVTQLREEAERRAQQQAEAERAREEAERELERWQLKANE 1780

Qy    271 SPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEK 330
Db   1781 ALRLRLQAEEVAQQKSL---------AQADAEKQKEEAEREARRRGKAEEQAVRQRELAE 1831

Qy    331 TEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEAS-SEEKCAPL---ATEVFDEK 386
Db   1832 QE-LEKQRQLTEGTAQQRLAAEQELIRLRAETEQGEHQRQLLEEELARLQHEATAATQKR 1890

Qy    387 MEAHQE---VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEA 443
Db   1891 QELEAELAKVRAEMEVLLASKARAEEESRSTSE----------------KSKQRLEAEA 1933

Qy    444 ----EPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE--GIVSE-VEMLSSQERIK 496
Db   1934 GRFRELAEEAARLRALAEEARRHRELA----EEDAARQRAEADGVLTEKLAAISEATRLK 1989

Qy    497 ----------VQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESA 545
Db   1990 TEAEIALKEKEAENERLRRLAEDEAFQRRRLEEQAAQHKADIEERLAQ---LRKASESEL 2046

Qy    546 DEQKG--ESSASSPEEPEETTCLEKGPLEAPQDGEAE--------EGTTSDGEKKREGIT 595
Db   2047 ERQKGLVEDTLRQRRQVEEEIMALKASFEKAAAGKAELELELGRIRSNAEDTMRSKELAE 2106

Qy    596 PWASFKKMVTPKKRVRRPSESDK-------EEELEKVKSATLSSTDSTVSEMQDEVKTVG 648
Db   2107 QEAARQRQLAAEEEQRRREAEERVQRSLAAEEEAARQRKVALEEVERLKAKV-EEARRLR 2165

Qy    649 EEQKPEEPK----------RRVDTSVSWEALIC-------------------VGSSKKR 678
Db   2166 ERAEQESARQLQLAQEAAQKRLQAEEKAHAFVVQQREEELQQTLQQEQNMLERLRSEAEA 2225

Qy    679 ARKASSSDDEG---GPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAG 735
Db   2226 ARRAAEEAEEAREQAEREAAQSRKQVEEAERLKQS------AEEQAQAQAQAQAAAE--- 2276

Qy    736 SPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIP 795
Db   2277 --------------------KLRKEAEQEAARRAQAEQAALKQKQAADAEMEKHKKFAE 2315

Qy    796 GRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKME-AQGNTELPQL 854
Db   2316 QTLRQKAQVEQELTTLR----LQLEETD-HQKSILD-EELQRLKAEVTEAARQRSQVEEE 2369

Qy    855 LGAVYVS-EELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEP-LEHTAGEAMPP 912
Db   2370 LFSVRVQMEELGK------------LKARIEAENRALILRDKDNTQRFLEEEAEKMKQV 2416

Qy    913 VEEVTEKDIIAEETPVLTQ----TLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEE 968
Db   2417 AEEAARLSVAAQEAARLRQLAEEDLAQQRALAEKMLKEKMQAVQEATRLKAEAELLQQQK 2476

Qy    969 VTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADK 1028
Db   2477 EL----AQEQARRLQA--------DKEQMAQQLVEETQGFQRTLEAERQRQLEMSAEAER 2524

Qy   1029 VK------EESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQ----E 1078
Db   2525 LKLRMAEMSRAQARAEEDAQRFRKQAEEIGEKLHRTE--LATQEKVTLVQTLEIQRQQSD 2582

Qy   1079 AGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQL-MEQAVAPESSETLTDS 1137
Db   2583 QDAERLREAIAELEREKEKLK---------------QEAKLLQLKSEEMQTVQQEQILQET 2628

Qy   1138 ETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVP 1197
Db   2629 QALQKSFLSEKDSL--LQRERFIEQEKAKL------EQLFQDEVAKAKQ----------- 2669

Qy   1198 AQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGP 1257
Db   2670 LQEE--QQRQQQQMEQEKQELVASMEEARRRQREAEEG---------VRRKQE------- 2711

Qy   1258 NSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQ 1317
Db   2712 -------------------------ELQRLEQQRQQQEKLLAEENQRLR-ERLQRLEEE 2744

Qy   1318 ETAAPEH--EGTYGKPVLT---------LDMPSSERGKALGSLGGSPSLPDQ--DKAGCI 1364
Db   2745 HRAALAHSEEIATSQAAATKALPNGRDALDGPSMEAEPEYTFEGLRQKVPAQQLQEAGIL 2804

Qy   1365 EV-QVQSLDTTVTQTAEAVEK-----------VIETVVISETGE-------------SPE 1399
```

```
Db   2805 SMEELQRLTQGHTTVAELTQREDVRHYLKGGSSIAGLLLKPTNEKLSVYTALQRQLLSP- 2863

Qy   1400 CVGAHLLPAEKSSATG--------GHWTLQHAEDTVPLGPE 1432
           |  |:  |  :|:|            |:  |     :|||
Db   2864 --GTALILLEAQAASGFLLDPVRNRRLTVNEAVKEGVVGPE 2902


RESULT 10
PLE1_CRIGR
ID   PLE1_CRIGR     STANDARD;      PRT;  4473 AA.
AC   Q9JI55;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Plectin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated
DE   protein) (IFAP300) (Fragment).
GN   PLEC1.
OS   Cricetulus griseus (Chinese hamster).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC   Cricetulus.
OX   NCBI_TaxID=10029;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20334248; PubMed=10873583;
RA   Clubb B.H., Chou Y.-H., Herrmann H., Svitkina T.M., Borisy G.G.,
RA   Goldman R.D.;
RT   "The 300-kDa intermediate filament-associated protein (IFAP300) is a
RT   hamster plectin ortholog.";
RL   Biochem. Biophys. Res. Commun. 273:183-187(2000).
RN   [2]
RP   PHOSPHORYLATION.
RX   MEDLINE=96215219; PubMed=8626512;
RA   Malecz N., Foisner R., Stadler C., Wiche G.;
RT   "Identification of plectin as a substrate of p34cdc2 kinase and
RT   mapping of a single phosphorylation site.";
RL   J. Biol. Chem. 271:8203-8208(1996).
CC   -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC       MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC       HEMIDESMOSOMES. MAY BE INVOLVED NOT ONLY IN THE CROSSLINKING AND
CC       STABILIZATION OF CYTOSKELETAL INTERMEDIATE FILAMENTS NETWORK, BUT
CC       ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC   -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC   -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC       VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
CC       AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC   -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC       INTERMEDIATE FILAMENTS DURING MITOSIS.
CC   -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 32 PLECTIN REPEATS.
CC   -!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF260753; AAF70372.1; -.
DR   HSSP; Q01082; 1BKR.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00681; Plectin; 20.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00250; PLEC; 32.
DR   SMART; SM00150; SPEC; 4.
DR   PROSITE; PS00019; ACTININ_1; PARTIAL.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
KW   Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW   Phosphorylation.
FT   NON_TER       1      1
FT   DOMAIN       <1   1259       GLOBULAR 1.
FT   DOMAIN     1260   2544       CENTRAL FIBROUS ROD DOMAIN.
FT   DOMAIN     2545   4473       GLOBULAR 2.
FT   DOMAIN       <1    192       ACTIN-BINDING.
FT   DOMAIN       <1     74       CH 1.
FT   DOMAIN       87    189       CH 2.
FT   REPEAT      449    508       SPECTRIN 1.
FT   REPEAT      529    613       SPECTRIN 2.
FT   REPEAT      626    719       SPECTRIN 3.
FT   REPEAT     1104   1204       SPECTRIN 4.
FT   DOMAIN     1258   2548       COILED COIL (POTENTIAL).
FT   REPEAT     2615   2652       PLECTIN 1.
FT   REPEAT     2653   2690       PLECTIN 2.
FT   REPEAT     2691   2728       PLECTIN 3.
FT   REPEAT     2729   2766       PLECTIN 4.
FT   REPEAT     2770   2804       PLECTIN 5.
FT   REPEAT     2905   2942       PLECTIN 6.
FT   REPEAT     2943   2980       PLECTIN 7.
FT   REPEAT     2981   3018       PLECTIN 8.
FT   REPEAT     3019   3056       PLECTIN 9.
FT   REPEAT     3057   3094       PLECTIN 10.
FT   REPEAT     3274   3311       PLECTIN 11.
FT   REPEAT     3312   3349       PLECTIN 12.
FT   REPEAT     3350   3387       PLECTIN 13.
FT   REPEAT     3388   3425       PLECTIN 14.
FT   REPEAT     3429   3463       PLECTIN 15.
FT   REPEAT     3609   3646       PLECTIN 16.
FT   REPEAT     3647   3684       PLECTIN 17.
FT   REPEAT     3685   3722       PLECTIN 18.
FT   REPEAT     3723   3760       PLECTIN 19.
FT   REPEAT     3764   3797       PLECTIN 20.
FT   REPEAT     3800   3834       PLECTIN 21.
FT   REPEAT     3852   3889       PLECTIN 22.
FT   REPEAT     3890   3927       PLECTIN 23.
FT   REPEAT     3928   3965       PLECTIN 24.
FT   REPEAT     3966   4003       PLECTIN 25.
FT   REPEAT     4007   4041       PLECTIN 26.
FT   REPEAT     4043   4094       PLECTIN 27.
FT   REPEAT     4197   4234       PLECTIN 28.
FT   REPEAT     4235   4272       PLECTIN 29.
FT   REPEAT     4273   4310       PLECTIN 30.
FT   REPEAT     4311   4348       PLECTIN 31.
FT   REPEAT     4349   4386       PLECTIN 32.
FT   DOMAIN     4039   4089       BINDING TO INTERMEDIATE FILAMENTS (BY
FT                                SIMILARITY).
FT   DOMAIN     4414   4429       4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT   MOD_RES    4328   4328       PHOSPHORYLATION (BY CDC2).
SQ   SEQUENCE   4473 AA;  509015 MW;  E144615D361E3484 CRC64;

  Query Match             5.0%;  Score 402.5;  DB 1;  Length 4473;
  Best Local Similarity   19.2%;  Pred. No. 1e-06;
  Matches  343;  Conservative  280;  Mismatches  774;  Indels  389;  Gaps   64;

Qy      8 EQRSPEQPAGSD----TPSELVLSGHGPAAEASGAAGDPADADPATKLPQK------NGQ 57
           |:|  ||     :          |    |      |||    |     |: :  | :|  ::          :
Db   1273 EERLAEQQRAEERERLAEGEAALEKQRQLAEAHAQAKAQAELE-AQELQRRMQEEVARRE 1331

Qy     58 LSSVNGVAEQGDVHVQEENQ---EGQEEEVVDE----DVGQRESEDVREKDRVEEMAANS 110
            ::||   |:|      :|||  |       :   |  |:      :   :|     : |:  ||    :     :
Db   1332 EAAVN--AQQQKRSIQEELQHLRQSSEAEIQAKAQQVEAAERSRMRIEEEIRVVRLQLET 1389

Qy    111 TAVEDITKDGQEETSEI-IEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVK- 168
           |   :     :|: :         |:   |  :      :|   :      |     |      :|              ||
Db   1390 TERQRGGAEGELQALRARAEEAEAQKRQAQEEAERLRRQVQDESQRKRQAEAELALRVKA 1449

Qy    169 --KDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQ---------EPSVETAVGESASKES 217
             :    ||     :|  |     :   :  |   ||     :           :              |  :   :|    |    ||  :
```

```
Db   1450 QAEAAQEKQRALQALEELRLQAEEAERRLRQAQAERARQVQVALETAQRSAEVELQSKRA 1509

Qy    218 ELKQSTEKQEGTLKQEQ-SSTEIPLQAE---SDQAAEEEAKDEGEEKQEKEPTKSPESPS 273

Db   1510 SFAEKTAQLERTLQEEHVTVTQLREKAERRAQQQAEAERAREEAERELERWQLKANEALR 1569

Qy    274 SPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEP 333

Db   1570 LRLQAEEVAQQRSL---------AQADAEKQKEEAEREARRRGKAEEQAVRQRELAEQE- 1619

Qy    334 ASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEAS-SEEKCAPL---ATEVFDEKMEA 389

Db   1620 LEKQRQLAEGTAQQRLAAEQELIRLRAETEQGEQQRQLLEEELARLQREATAATHKRQEL 1679

Qy    390 HQEVV-----AEVHVSTVEKTEEEQGGGE---------------AEGGVVVEGTGESL 428

Db   1680 EAELAKVRAEMEVLLASKARAEEESRSTSEKSKQRLEAEADRFRELAEEAARLRALAEEA 1739

Qy    429 PPEK-LAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVE 487

Db   1740 KRQRQLAEEDAARQRAE-AERVLTEKLAAIS-----EATRLKTEAEIALKEKEAENERLR 1793

Qy    488 MLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADE 547

Db   1794 RLAEDEAFQ-----RRRLEEQAALHKADIEERLAQ--------------LRKASESELER 1834

Qy    548 QKG--ESSASSPEEPEETTCLEKGPLEAPQDGEAE--------EGTTSDGEKKREGITPW 597

Db   1835 QKGLVEDTLRQRRQVEEEILALKVSFEKAAAGKAELELELGRIRSSAEDTMRSKEQAEQE 1894

Qy    598 ASFKKMVTPKKRVRRPSESDK-------EEELEKVKSATLSSTDSTVSEMQDEVKTVGEE 650

Db   1895 AARQRQLAAEEEQRRREAEERVQKSLAAEEEAARQRKAALEEVERLKAKV-EEARRLRER 1953

Qy    651 QKPEEPK----------RRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHR 700

Db   1954 AEQESARQLQLAQEAAQKRLQAEEKAHAFVV--QQREEELQQTLQQEQSMLERLRGEAEA 2011

Qy    701 AEEASKD-KEAGTDAVPASTQEQDQAQGS-----SSPEPAGSPSEGEGVSTWESFKRLVT 754

Db   2012 ARRAAEEAEEAREQAEREAAQSRKQVEEAERLKQSAEEQAQARAQAQAAA---------- 2061

Qy    755 PRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDS 814

Db   2062 -EKLRKEAEQEAARRAQAEQAALKQKQAADAEMEKHKKFAEQTLRQKAQVEQELTTLR-- 2118

Qy    815 GPVEINEDDPNVPAVVPLSEYNAVEREKME-AQGNTELPQLLGAVYVS-EELSKTLVHTV 872

Db   2119 --LQLEETD-HQKSILD-EELQRLKAEVTEAARQRSQVEEELFSVRVQMEELGK------ 2168

Qy    873 SVAVIDGTRAVTSVEERSPSWISASVTEP-LEHTAGEAMPPVEEVTEKDIIAEETPVLTQ 931

Db   2169 -------LKARIEAENRALILRDKDNTQRFLEEEAEKMKQVAEEAARLSVAAQEAARLRQ 2221

Qy    932 ----TLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQ 987

Db   2222 LAEEDLAQQRALAEKMLKEKMQAVQEATRLKAEAELLQQQKELAQEQARRLQE------- 2274

Qy    988 LTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK--------------EES 1033

Db   2275 -----DKEQMAQQLVEETQGFQRTLEVERQRQLEMSAEAERLKLRMAEMSRAQARAEEDA 2329

Qy   1034 Q----------------------VPATQT--VQRTGS-----KALEKVEEVEEDSE--- 1060

Db   2330 QRFRKQAEEIGEKLHRTELATQEKVTLVQTLEIQRQQSDHDAERLREAIAELEREKEKLK 2389

Qy   1061 ------VLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQ 1114

Db   2390 QEAKLLQLKSEEMQTVQQEQILQETQALQKSFLSEKDSLLQRERFIEQEKAKLEQLFQDE 2449

Qy   1115 VIKLQQLME-----QAVAPESSETLTDSETNGSTPLADSDTADGTQQDE----TIDSQDS 1165

Db   2450 VAKAQQLREEQQRQQRQMEQEKQELVASMEEARRRQCEAEEAVRRKQEELQHLELQRQQQ 2509

Qy   1166 KATAAVRQSQVTE----------------EEAATAQKEEPSTLPNNVPAQEEHGEEPGRD 1209

Db   2510 EKLLAEENQRLRERLQRLEEEHRAALAHSEEIAATQAAAAKALPN------------GRD 2557

Qy   1210 VLEPTQQEL--------TAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQK 1261

Db   2558 ALDGPSMEVEPEHAFEGLRQKVPALQLREAG------ILSAEELQRLEQGHTTVAELSQR 2611

Qy   1262 AADVTYDSEVMGVAGCQEKESTEVQS---------------LSLEEGEMET----DVEKE 1302

Db   2612 EDVRQYLQGRSSIAGLLLKPTDEKLSVYTALQRQLLSPGTALILLEAQAASGFLLDPVRN 2671

Qy   1303 KRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSER---GKALGSLGGSPSLPDQD 1359

Db   2672 RRLTVNE------------PVKEGVVG-PELHHKLLSAERAVTGYKDPYTGEQISLFQAM 2718

Qy   1360 KAGCIEVQVQSLDTTVTQTAEAVEKVIETV----------------------VISETG-- 1395

Db   2719 KK---DLLVRDHAIRLLEAQIATGGIIDTVHSHRVPVDVAYQRGYFDEEMSRILADPGDD 2775

Qy   1396 ----------------------ESPECVGAHLLPAEKSSATGGH--WTLQHAEDTVPLG 1430

Db   2776 TKGFFDPNTHENLTYLQLLERCVEDPE-TGLHLLPLTDKAAKGGELVYTDTEARDVF--- 2831

Qy   1431 PESQAESIPIIVTPAPESTLHPDLQGE-ISASQRERSEEEDKPDAGPDADGKESTAIEKV 1489

Db   2832 -EKATVSAPFGKFQGKTVTIWEIINSEYFTAEQRRDLLRQFRTG----------RITVEKI 2881

Qy   1490 LKAEPEILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQ 1535

Db   2882 IKIVITVVEEQERKGQLCFEGLRALVPAAELLESGVISHELYQQLQ 2927
```

RESULT 11
PLE1_HUMAN
ID   PLE1_HUMAN   STANDARD;   PRT; 4684 AA.
AC   Q15149; Q16640; Q15148;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Plectin 1 (PLTN) (PCN) (Hemidesmosomal protein 1) (HD1).
GN   PLEC1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Placenta;
RX   MEDLINE=96210632; PubMed=8633055;
RA   Liu C.-G., Maercker C., Castanon M.J., Hauptmann R., Wiche G.;
RT   "Human plectin: organization of the gene, sequence analysis, and
RT   chromosome localization (8q24).";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:4278-4283(1996).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), AND DISEASE.
RX   MEDLINE=96312447; PubMed=8698233;
RA   McLean W.H.I., Pulkkinen L., Smith F.J.D., Rugg E.L., Lane E.B.,
RA   Bullrich F., Burgeson R.E., Amano S., Hudson D.L., Owaribe K.,
RA   McGrath J.A., McMillan J.R., Eady R.A.J., Leigh I.M., Christiano A.M.,
RA   Uitto J.;
RT   "Loss of plectin causes epidermolysis bullosa with muscular dystrophy:
RT   cDNA cloning and genomic organization.";
RL   Genes Dev. 10:1724-1735(1996).
RN   [3]
RP   VARIANT MD-EBS 1003-GLN--ALA-1005 DEL.
RX   MEDLINE=97049959; PubMed=8894687;
RA   Pulkkinen L., Smith F.J.D., Shimizu H., Murata S., Yaoita H.,
RA   Hachisuka H., Nishikawa T., McLean W.H.I., Uitto J.;
RT   "Homozygous deletion mutations in the plectin gene (PLEC1) in patients
RT   with epidermolysis bullosa simplex associated with late-onset
RT   muscular dystrophy.";

```
RL   Hum. Mol. Genet. 5:1539-1546(1996).
RN   [4]
RP   VARIANT MD-EBS LEU-429 INS.
RX   MEDLINE=21090821; PubMed=11159198;
RA   Bauer J.W., Rouan F., Kofler B., Rezniczek G.A., Kornacker I.,
RA   Muss W., Hametner R., Klausegger A., Huber A., Pohla-Gubo G.,
RA   Wiche G., Uitto J., Hintner H.;
RT   "A compound heterozygous one amino-acid insertion/nonsense mutation in
RT   the plectin gene causes epidermolysis bullosa simplex with plectin
RT   deficiency.";
RL   Am. J. Pathol. 158:617-625(2001).
CC   -!- FUNCTION: INTERLINKS INTERMEDIATE FILAMENTS WITH MICROTUBULES AND
CC       MICROFILAMENTS AND ANCHORS INTERMEDIATE FILAMENTS TO DESMOSOMES OR
CC       HEMIDESMOSOMES. COULD ALSO BIND MUSCLE PROTEINS SUCH AS ACTIN TO
CC       MEMBRANE COMPLEXES IN MUSCLE. MAY BE INVOLVED NOT ONLY IN THE
CC       CROSSLINKING AND STABILIZATION OF CYTOSKELETAL INTERMEDIATE
CC       FILAMENTS NETWORK, BUT ALSO IN THE REGULATION OF THEIR DYNAMICS.
CC   -!- SUBUNIT: HOMODIMER OR HOMOTETRAMER.
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
CC   -!- TISSUE SPECIFICITY: WIDELY EXPRESSED WITH HIGHEST LEVELS IN
CC       MUSCLE, HEART, PLACENTA AND SPINAL CORD.
CC   -!- DOMAIN: THE N-TERMINUS INTERACTS WITH ACTIN, THE C-TERMINUS WITH
CC       VIMENTIN, DESMIN, GFAP, CYTOKERATINS, LAMIN B; WHEREAS BOTH THE N-
CC       AND THE C-TERMINUS CAN BIND INTEGRIN BETA-4.
CC   -!- PTM: PHOSPHORYLATED BY CDC2; REGULATES DISSOCIATION FROM
CC       INTERMEDIATE FILAMENTS DURING MITOSIS (BY SIMILARITY).
CC   -!- DISEASE: DEFECTS IN PLEC1 ARE THE CAUSE OF EPIDERMOLYSIS BULLOSA
CC       WITH MUSCULAR DYSTROPHY (MD-EBS OR EB-MD), AN AUTOSOMAL RECESSIVE
CC       DISORDER CHARACTERIZED BY EPIDERMAL BLISTER FORMATION AT THE LEVEL
CC       OF THE HEMIDESMOSOME AND ASSOCIATED LATE-ONSET MUSCULAR DYSTROPHY.
CC   -!- SIMILARITY: CONTAINS 1 ACTIN-BINDING DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC   -!- SIMILARITY: CONTAINS 33 PLECTIN REPEATS.
CC   -!- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
CC   -!- SIMILARITY: BELONGS TO THE PLAKIN OR CYTOLINKER FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z54367; CAA91196.1; -.
DR   EMBL; U53204; AAB05427.1; -.
DR   EMBL; U63610; AAB05428.1; -.
DR   EMBL; U63609; AAB05428.1; JOINED.
DR   EMBL; X97053; CAA65765.1; -.
DR   HSSP; Q01082; 1BKR.
DR   Genew; HGNC:9069; PLEC1.
DR   MIM; 601282; -.
DR   MIM; 226670; -.
DR   InterPro; IPR001589; Actbind_actnin.
DR   InterPro; IPR001715; Calponin-like.
DR   InterPro; IPR001101; Plectin_repeat.
DR   InterPro; IPR005326; S10_plectin.
DR   InterPro; IPR002017; Spectrin.
DR   Pfam; PF00307; CH; 2.
DR   Pfam; PF00681; Plectin; 19.
DR   Pfam; PF03501; S10_plectin; 1.
DR   SMART; SM00033; CH; 2.
DR   SMART; SM00250; PLEC; 33.
DR   SMART; SM00150; SPEC; 5.
DR   PROSITE; PS00019; ACTININ_1; FALSE_NEG.
DR   PROSITE; PS00020; ACTININ_2; FALSE_NEG.
DR   PROSITE; PS50021; CH; 2.
KW   Coiled coil; Repeat; Structural protein; Cytoskeleton; Actin-binding;
KW   Phosphorylation; Alternative splicing; Epidermolysis bullosa;
KW   Disease mutation.
FT   DOMAIN        1   1470       GLOBULAR 1.
FT   DOMAIN     1471   2755       CENTRAL FIBROUS ROD DOMAIN.
```

```
FT   DOMAIN     2756   4684       GLOBULAR 2.
FT   DOMAIN      175    400       ACTIN-BINDING.
FT   DOMAIN      179    282       CH 1.
FT   DOMAIN      295    397       CH 2.
FT   REPEAT      645    710       SPECTRIN 1.
FT   REPEAT      740    824       SPECTRIN 2.
FT   REPEAT      837    930       SPECTRIN 3.
FT   REPEAT     1315   1415       SPECTRIN 4.
FT   DOMAIN     1469   2756       COILED COIL (POTENTIAL).
FT   REPEAT     2826   2863       PLECTIN 1.
FT   REPEAT     2864   2901       PLECTIN 2.
FT   REPEAT     2902   2939       PLECTIN 3.
FT   REPEAT     2940   2977       PLECTIN 4.
FT   REPEAT     2981   3015       PLECTIN 5.
FT   REPEAT     3116   3153       PLECTIN 6.
FT   REPEAT     3154   3191       PLECTIN 7.
FT   REPEAT     3192   3229       PLECTIN 8.
FT   REPEAT     3230   3267       PLECTIN 9.
FT   REPEAT     3268   3305       PLECTIN 10.
FT   REPEAT     3306   3343       PLECTIN 11.
FT   REPEAT     3485   3522       PLECTIN 12.
FT   REPEAT     3523   3560       PLECTIN 13.
FT   REPEAT     3561   3598       PLECTIN 14.
FT   REPEAT     3599   3636       PLECTIN 15.
FT   REPEAT     3640   3674       PLECTIN 16.
FT   REPEAT     3820   3857       PLECTIN 17.
FT   REPEAT     3858   3895       PLECTIN 18.
FT   REPEAT     3896   3933       PLECTIN 19.
FT   REPEAT     3934   3971       PLECTIN 20.
FT   REPEAT     3975   4008       PLECTIN 21.
FT   REPEAT     4063   4100       PLECTIN 22.
FT   REPEAT     4101   4138       PLECTIN 23.
FT   REPEAT     4139   4176       PLECTIN 24.
FT   REPEAT     4177   4214       PLECTIN 25.
FT   REPEAT     4218   4252       PLECTIN 26.
FT   REPEAT     4265   4305       PLECTIN 27.
FT   REPEAT     4319   4356       PLECTIN 28.
FT   REPEAT     4408   4445       PLECTIN 29.
FT   REPEAT     4446   4483       PLECTIN 30.
FT   REPEAT     4484   4521       PLECTIN 31.
FT   REPEAT     4522   4559       PLECTIN 32.
FT   REPEAT     4560   4597       PLECTIN 33.
FT   DOMAIN     4250   4300       BINDING TO INTERMEDIATE FILAMENTS (BY
FT                                SIMILARITY).
FT   DOMAIN     4625   4640       4 X 4 AA TANDEM REPEATS OF G-S-R-X.
FT   MOD_RES    4539   4539       PHOSPHORYLATION (BY CDC2) (BY
FT                                SIMILARITY).
FT   VARSPLIC      1    174       MVAGMLMPRDQLRAIYEVLFREGVMVAKKDRRPRSLHPHVP
FT                                GVTNLQVMRAMASLRARGLVRETFAWCHFFWYLTNEGIAHL
FT                                RQYLHLPPEIVAASLQRVRRPVAMVMPARRTPHVQAVQGPL
FT                                GSPPKRGPLPTEEQRLYRRKELEEVSPETPVVPATTQRTLA
FT                                RPGPEPAPAT -> MSGEDAEVRAVSEDVSNGSSGSPSPGD
FT                                TLPWNLGKTQRSRRSGGGAGSNGSVLDPAERAVIRIA (IN
FT                                ISOFORM 2 AND ISOFORM 3).
FT   VARSPLIC    409    412       MISSING (IN ISOFORM 3).
FT   VARIANT     429    429       L -> LL (IN MD-EBS).
FT                                /FTId=VAR_011336.
FT   VARIANT    1003   1005       MISSING (IN MD-EBS).
FT                                /FTId=VAR_011337.
FT   CONFLICT    185    185       S -> F (IN REF. 2).
FT   CONFLICT    259    259       N -> D (IN REF. 2).
FT   CONFLICT    550    550       H -> N (IN REF. 2).
FT   CONFLICT    560    560       I -> V (IN REF. 2).
FT   CONFLICT    706    706       Q -> R (IN REF. 2).
FT   CONFLICT    886    886       N -> Y (IN REF. 2).
FT   CONFLICT   1002   1002       V -> A (IN REF. 2).
FT   CONFLICT   1309   1309       V -> L (IN REF. 2).
FT   CONFLICT   1321   1321       L -> V (IN REF. 2).
FT   CONFLICT   1334   1334       V -> L (IN REF. 2).
FT   CONFLICT   1534   1534       I -> M (IN REF. 2).
FT   CONFLICT   1662   1662       A -> T (IN REF. 2).
FT   CONFLICT   1688   1690       WLC -> RLR (IN REF. 2).
FT   CONFLICT   1767   1767       Q -> E (IN REF. 2).
```

```
FT   CONFLICT   1789   1789      L -> A (IN REF. 2).
FT   CONFLICT   1910   1910      K -> R (IN REF. 2).
FT   CONFLICT   2154   2154      K -> N (IN REF. 2).
FT   CONFLICT   2160   2160      S -> R (IN REF. 2).
FT   CONFLICT   2215   2215      R -> Q (IN REF. 2).
FT   CONFLICT   2244   2244      S -> A (IN REF. 2).
FT   CONFLICT   3027   3027      K -> E (IN REF. 2).
FT   CONFLICT   3310   3310      E -> A (IN REF. 2).
FT   CONFLICT   3361   3361      F -> L (IN REF. 2).
FT   CONFLICT   3408   3408      F -> L (IN REF. 2).
FT   CONFLICT   3447   3447      S -> A (IN REF. 2).
FT   CONFLICT   3531   3531      G -> A (IN REF. 2).
FT   CONFLICT   3580   3580      R -> S (IN REF. 2).
FT   CONFLICT   3589   3589      K -> Q (IN REF. 2).
FT   CONFLICT   3596   3596      E -> Q (IN REF. 2).
FT   CONFLICT   3616   3616      N -> H (IN REF. 2).
FT   CONFLICT   3686   3686      V -> A (IN REF. 2).

  Query Match             5.0%;  Score 400.5;  DB 1;  Length 4684;
  Best Local Similarity   19.7%;  Pred. No. 1.2e-06;
  Matches  313;  Conservative  275;  Mismatches  667;  Indels  331;  Gaps   62;

Qy      8 EQRSPEQPAGSD----TPSELVLSGHGPAAEASGAAGDPADADPATKLPQK------NGQ 57
Db   1484 EERLAEQQRAEERERLAEVEAALEKQRQLAEAHAQAKAQAERE-AKELQQRIQEEVVRRE 1542

Qy     58 LSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESE-------DVREKDRVEEMAANS 110
Db   1543 EAAVD--AQQQKRSIQEELQQLRQSSEAEIQAKARQAEAAERSRLRIEEEIRVVRLQLEA 1600

Qy    111 TAVEDITKDGQEETSEI-IEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKK 169
Db   1601 TERQRGGAEGELQALRARAEEAEAQKRQAQEEAERLRRQVQDESQRKRQAEVELASRVKA 1660

Qy    170 D---KNEKSDTVQLLTVKKDEGEGAE-----ASVGAGDHQEPSVET----AVGESASKES 217
Db   1661 EAEAAREKQRALQALEELRLQAEEAERWLCQAEVERARQVQVALETAQRSAEAELQSKRA 1720

Qy    218 ELKQSTEKQEGTLKQEQ-SSTEIPLQAE---SDQAAEEEAKDEGEEKQEKEPTKSPESPS 273
Db   1721 SFAEKTAQLERSLQEEHVAVAQLREEAERRAQQQAEAERAREEAERQLERWQLKANEALR 1780

Qy    274 SPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEP 333
Db   1781 LRLQAEEVLQQKSL---------AQAEAEKQKEEAEREARRRGKAEEQAVRQRELAEQE- 1830

Qy    334 ASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEAS-SEEKCAPL---ATEVFDEKMEA 389
Db   1831 LEKQRQLAEGTAQQRLAAEQELIRLRAETEQGEQQRQLLEEELARLQREAAAATQKRQEL 1890

Qy    390 HQEVV-----AEVHVSTVEKTEEEQGGGG-------EAEGGVVVEGTGE-----SLPPE- 431
Db   1891 EAELAKVRAEMEVLLASKAKAEEESRSTSEKSKQRLEAEAGRFRELAEEAARLRALAEEA 1950

Qy    432 ----KLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVE 487
Db   1951 KRQRQLAEEDAARQRAE-AERVLAEKLAAI--GEATRL-------KT-----EAEIALKE 1995

Qy    488 MLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADE 547
Db   1996 KEAENER-------LRRLAEDEAFQRRRLEEQAAQHKADIEERLAQ---LRKASDSELER 2045

Qy    548 QKG--ESSASSPEEPEETTCLEKGPLEAPQDGEAE--------EGTTSDGEKKREGITPW 597
Db   2046 QKGLVEDTLRQRRQVEEEILALKASFEKAAAGKAELELELGRIRSNAEDTLRSKEQAELE 2105

Qy    598 ASFKKMVTPKKRVRRPSESDK-------EEELEKVKSATLSSTDSTVSEMQDEVKTVGEE 650
Db   2106 AARQRQLAAEEERRRREAEERVQKSLAAEEEAARQRKAALEEVERLKAKV-EEARSLRER 2164

Qy    651 QKPEEPK----------RRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHR 700
Db   2165 AEQESARQLQLAQEAAQKRLQAEEKAHAFAV--QQKEQELQQTLQQEQSVLDRLRGEAEA 2222

Qy    701 AEEASKD-KEAGTDAVPASTQEQDQAQGS-----SSPEPAGSPSEGEGVSTWESFKRLVT 754
Db   2223 ARRAAEEAEEARVQAEREAAQSRRQVEEAERLKQSAEEQAQARAQAQAAA---------- 2272

Qy    755 PRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDS 814
Db   2273 -EKLRKEAEQEAARRAQAEQAALRQKQAADAEMEKHKKFAEQTLRQKAQVEQELTTLR-- 2329

Qy    815 GPVEINEDDPNVPAVVPLSEYNAVERE----KME----AQGNTELPQLLGAVYVS-EELS 865
Db   2330 --LQLEETD---------HQKNLLDEELQRLKAEATEAARQRSQVEEELFSVRVQMEELS 2378

Qy    866 KTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPV-EEVTEKDIIAE 924
Db   2379 K------------LKARIEAENRALILRDKDNTQRFLQEEAEKMKQVAEEAARLSVAAQ 2425

Qy    925 ETPVLTQ----TLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTD 980
Db   2426 EAARLRQLAEEDLAQQRALAEKMLKEKMQAVQEATRLKAEAELLQQQKELAQEQARRLQE 2485

Qy    981 MVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK------EESQ 1034
Db   2486 ------------DKEQMAQQLAEETQGFQRTLEAERQRQLEMSAEAERLKLRVAEMSRAQ 2533

Qy   1035 VPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGSETGQAT 1094
Db   2534 ARAEEDAQRFRKQAEEIGEKLHRTE--LATQEKVTLVQTLEIQRQQSDHDAE-------- 2583

Qy   1095 PESLEVPEVTADVDH--VATCQVIKLQQL-MEQAVAPESSETLTDSETNGSTPLADSDTA 1151
Db   2584 ----RLREAIAELEREKEKLQQEAKLLQLKSEEMQTVQQEQLLQETQALQQSFLSEKDSL 2639

Qy   1152 DGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVL 1211
Db   2640 --LQRERFIEQEKAKL------EQLFQDEVAKAQQ-----------LREE--QQRQQQQM 2678

Qy   1212 EPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEV 1271
Db   2679 EQERQRLVASMEEARRRQHEAEEG---------VRRKQE--------------------- 2708

Qy   1272 MGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPE-QVSEEGEQETAAPEHEGTYGK 1330
Db   2709 ------------ELQQLEQQRRQQEELLAEENQRLREQLQLLEEQHRAALAHSEEVTASQ 2756

Qy   1331 PVLTLDMPS-----------SERGKALGSLGGSPSLPDQDKAGCIEV-QVQSL---DTTV 1375
Db   2757 VAATKTLPNGRDALDGPAAEAEPEHSFDGLRRKVSAQRLQEAGILSAEELQRLAQGHTTV 2816

Qy   1376 TQTAEAVE--------KVIETVVISETGE------------SPECVGAHLLPAEKSSAT 1414
Db   2817 DELARREDVRHYLQGRSSIAGLLLKATNEKLSVYAALQRQLLSP---GTALILLEAQAAS 2873

Qy   1415 G--------GHWTLQHAEDTVPLGPE 1432
Db   2874 GFLLDPVRNRRLTVNEAVKEGVVGPE 2899

RESULT 12
PGCV_HUMAN
ID   PGCV_HUMAN      STANDARD;      PRT;  3396 AA.
AC   P13611; P20754; Q9UNW5; Q13010; Q13189; Q15123;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Versican core protein precursor (Large fibroblast proteoglycan)
DE   (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
DE   hyaluronate-binding protein) (GHAP).
GN   CSPG2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
```

```
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM V0).
RX   MEDLINE=95105188; PubMed=7528742;
RA   Naso M.F., Zimmermann D.R., Iozzo R.V.;
RT   "Characterization of the complete genomic structure of the human
RT   versican gene and functional analysis of its promoter.";
RL   J. Biol. Chem. 269:32999-33008(1994).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM V1).
RC   TISSUE=Placenta;
RX   MEDLINE=90059882; PubMed=2583089;
RA   Zimmermann D.R., Ruoslahti E.;
RT   "Multiple domains of the large fibroblast proteoglycan, versican.";
RL   EMBO J. 8:2975-2981(1989).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM V2).
RC   TISSUE=Glial tumor;
RX   MEDLINE=95105187; PubMed=7806529;
RA   Dours-Zimmermann M.T., Zimmermann D.R.;
RT   "A novel glycosaminoglycan attachment domain identified in two
RT   alternative splice variants of human versican.";
RL   J. Biol. Chem. 269:32992-32998(1994).
RN   [4]
RP   SEQUENCE OF 2711-3396 FROM N.A.
RC   TISSUE=Lung fibroblast;
RX   MEDLINE=88007514; PubMed=2820964;
RA   Krusius T., Gehlsen K.R., Ruoslahti E.;
RT   "A fibroblast chondroitin sulfate proteoglycan core protein contains
RT   lectin-like and growth factor-like sequences.";
RL   J. Biol. Chem. 262:13120-13125(1987).
RN   [5]
RP   SEQUENCE OF 251-347 FROM N.A.
RX   MEDLINE=93122792; PubMed=1478664;
RA   Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
RA   McPherson J.D.;
RT   "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
RT   human chromosome 5 (5q12-5q14).";
RL   Genomics 14:845-851(1992).
RN   [6]
RP   SEQUENCE FROM N.A. (ISOFORM V3).
RC   TISSUE=Brain;
RX   MEDLINE=95181355; PubMed=7876137;
RA   Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT   "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT   without a chondroitin sulfate attachment in region in mouse and human
RT   tissues.";
RL   J. Biol. Chem. 270:3914-3918(1995).
RN   [7]
RP   SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
RC   TISSUE=Aortic smooth muscle;
RX   MEDLINE=99327053; PubMed=10397680;
RA   Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
RA   Wight T.N.;
RT   "Versican/PG-M isoforms in vascular smooth muscle cells.";
RL   Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
RN   [8]
RP   PARTIAL SEQUENCE.
RC   TISSUE=Brain;
RX   MEDLINE=89174663; PubMed=2466833;
RA   Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
RT   "Isolation and partial characterization of a glial
RT   hyaluronate-binding protein.";
RL   J. Biol. Chem. 264:5981-5987(1989).
RN   [9]
RP   TISSUE SPECIFICITY OF ISOFORMS.
RX   MEDLINE=96213482; PubMed=8627343;
RA   Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
RT   "Differential expression of versican isoforms in brain tumors.";
RL   J. Neuropathol. Exp. Neurol. 55:528-533(1996).
CC   -!- FUNCTION: May play a role in intercellular signaling and in
CC       connecting cells with the extracellular matrix. May take part in
CC       the regulation of cell motility, growth and differentiation. Binds
CC       hyaluronic acid.
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC   -!- ALTERNATIVE PRODUCTS: At least 5 isoforms; V0 (shown here), V1,
CC       V2, V3 and Vint; are produced by alternative splicing.
CC   -!- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
CC       in normal brain, gliomas, medulloblastomas, schwannomas,
CC       neurofibromas, and menningiomas; v2 is restricted to normal brain
CC       and gliomas; v3 is found in all these tissues except
CC       medulloblastomas.
CC   -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC   -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC   -!- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC   -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC   -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC   -!- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC   -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U16306; AAA65018.1; -.
DR   EMBL; X15998; CAA34128.1; -.
DR   EMBL; S52488; AAB24878.1; -.
DR   EMBL; U26555; AAA67565.1; -.
DR   EMBL; D32039; BAA06801.1; -.
DR   EMBL; J02814; AAA36437.1; -.
DR   EMBL; AF084545; AAD48545.1; -.
DR   PIR; S06014; S06014.
DR   PIR; A29348; A29348.
DR   PIR; A30358; A30358.
DR   HSSP; P01132; 1EGF.
DR   Genew; HGNC:2464; CSPG2.
DR   MIM; 118661; -.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000561; EGF-like.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR001438; EGF_II.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 2.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_c; 1.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 2.
DR   PRINTS; PR00010; EGFBLOOD.
DR   ProDom; PD000918; Link; 2.
DR   SMART; SM00032; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00001; EGF_like; 1.
DR   SMART; SM00409; IG; 1.
DR   SMART; SM00445; LINK; 2.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00022; EGF_1; 2.
DR   PROSITE; PS01186; EGF_2; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS01241; LINK; 2.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
KW   Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW   Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW   Hyaluronic acid; Alternative splicing.
FT   SIGNAL        1     20       POTENTIAL.
FT   CHAIN        21   3396       VERSICAN CORE PROTEIN.
FT   DOMAIN       37    137       IG-LIKE V-TYPE DOMAIN.
```

```
FT   DOMAIN      167    244       LINK 1.
FT   DOMAIN      265    346       LINK 2.
FT   DOMAIN      348   1335       GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
FT                                DOMAIN).
FT   DOMAIN     1336   3089       GAG-BETA.
FT   DOMAIN     3089   3125       EGF-LIKE 1.
FT   DOMAIN     3127   3163       EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     3176   3290       C-TYPE LECTIN.
FT   DOMAIN     3295   3353       SUSHI.
FT   DISULFID     44    130       BY SIMILARITY.
FT   DISULFID    172    243       BY SIMILARITY.
FT   DISULFID    196    217       BY SIMILARITY.
FT   DISULFID    270    345       BY SIMILARITY.
FT   DISULFID    294    315       BY SIMILARITY.
FT   DISULFID   3093   3104       BY SIMILARITY.
FT   DISULFID   3098   3113       BY SIMILARITY.
FT   DISULFID   3115   3124       BY SIMILARITY.
FT   DISULFID   3131   3142       BY SIMILARITY.
FT   DISULFID   3136   3151       BY SIMILARITY.
FT   DISULFID   3153   3162       BY SIMILARITY.
FT   DISULFID   3169   3180       BY SIMILARITY.
FT   DISULFID   3197   3289       BY SIMILARITY.
FT   DISULFID   3265   3281       BY SIMILARITY.
FT   DISULFID   3296   3339       BY SIMILARITY.
FT   DISULFID   3325   3352       BY SIMILARITY.
FT   CARBOHYD     57     57       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    615    615       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    782    782       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    809    809       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1332   1332       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1398   1398       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1442   1442       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1468   1468       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1663   1663       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   1898   1898       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2179   2179       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2272   2272       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2280   2280       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2360   2360       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD   2385   2385       N-LINKED (GLCNAC. . .) (POTENTIAL).

  Query Match             4.9%;  Score 395;  DB 1;  Length 3396;
  Best Local Similarity   19.9%;  Pred. No. 1.4e-06;
  Matches  431;  Conservative  252;  Mismatches  725;  Indels  756;  Gaps  102;

Qy         25 VLSGHGP-AAEASGAAGDPADADPAT-------------KLPQ-------KNGQLSSVNG 63
Db        786 VLLAHGTLSVEAATVSKWSWDEDNTTSKPLESTEPSASSKLPPALLTTVGMNGKDKDIPS 845

Qy         64 VAEQG--DVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQ 121
Db        846 FTEDGADEFTLIPDSTQKQLEEVTDEDIAAHGKFTIRFQPTTSTGIAEKSTLRDST---- 901

Qy        122 EETSEIIEQIPASENNV-EEMVQPAESQANDVGFKKVFKFV----------GFKFTVKKD 170
Db        902 --TEEKVPPITSTEGQVYATMEGSALGEVEDVDLSKPVSTVPQFAHTSEVEGLAFVSYSS 959

Qy        171 KNE------KSDTVQLLTVKKDEGEGAEASVGAGDH--QEPSVETAVGESASKES----E 218
Db        960 TQEPTTYVDSSHTIPLSVIPKTDWGVLVPSVPSEDEVLGEPSQDILVIDQTRLEATISPE 1019

Qy        219 LKQSTEKQEGTLKQE----QSSTEIPLQA-----------------ESDQAAEEEAKDE 256
Db       1020 TMRTTKITEGTTQEEFPWKEQTAEKPVPALSSTAWTPKEAVTPLDEQEGDGSAYTVSEDE 1079

Qy        257 ---GEE----------------------------KQEKEPTKSPE-------SPSSPVN 277
Db       1080 LLTGSERVPVLETTPVGKIDHSVSYPPGAVTEHKVKTDEVVTLTPRIGPKVSLSPGPEQK 1139

Qy        278 SET--------TSSFKKFFTHGWAGWRKKTSFKKSKED-DLETAEKRKEQEAEKVDEEEK 328
Db       1140 YETEGSSTTGFTSSLSPFSTHITQLMEETTTEKTSLEDIDLGSGLFEKPKATELIEFSTI 1199

Qy        329 EKTEPAS----------------------------EEQEPAEDTDQARLSADYEKVELP 359
Db       1200 KVTVPSDITTAFSSVDRLHTTSAFKPSSAITKKPPLIDREPGEETTSDMVIIGESTSHVP 1259

Qy        360 ---LEDQVG-----DLEASSEEKCAPLATE------VFDEKMEAHQEVVA---------- 395
Db       1260 PTTLEDIVAKETETDIDREYFTTSSPPATQPTRPPTVEDKEAFGPQALSTPQPPASTKFH 1319

Qy        396 -EVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELM---- 450
Db       1320 PDINVYIIEVRENKTGRMSD------LSVIGHPIDSES-KEDEPCSEETDPVHDLMAEIL 1372

Qy        451 -----------------KSREMCVSGGDHT-----QLTDLSPEEKTLPKHPEG------ 481
Db       1373 PEFPDIIEIDLYHSEENEEEEEECANATDVTTTPSVQYINGKHLVTTVPKDPEAAEARRG 1432

Qy        482 ---------------------IVSEVEM---------LSSQERIKVQGS---------- 500
Db       1433 QFESVAPSQNFSDSSESDTHPFVIAKTELSTAVQPNESTETTESLEVTWKPETYPETSEH 1492

Qy        501 ------------PLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE-------- 540
Db       1493 FSGGEPDVFPTVPFHEEFESGTAKK--GAESVTER----DTEVGHQAHEHTEPVSLFPEE 1546

Qy        541 -SPESADEQKGESSA-------SSPEEPEETTCLEKGPLEAPQDGEA----EEGTTSDGE 588
Db       1547 SSGEIAIDQESQKIAFARATEVTFGEEVEKSTSVTYTPTIVPSSASAYVSEEEAVTLIG- 1605

Qy        589 KKREGITPW----ASFKKM---VTPKKRV-----------RRPSESDKEEELEKVKSATL 630
Db       1606 ------NPWPDDLLSTKESWVEATPRQVVELSGSSSIPITEGSGEAEEDEDTMFTMVTDL 1659

Qy        631 SSTDST-----------VSEMQDEVKTVGEEQKPEEPKRRV---------DTSVSWEALI 670
Db       1660 SQRNTTDTLITLDTSRIITESFFEVPATTIYPVSEQPSAKVVPTKFVSETDTS-EWISST 1718

Qy        671 CVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQ---AQG 727
Db       1719 TVEEKKRK-------EEEGTTGT---------------ASTFEVYSSTQRSDQLILPFE 1755

Qy        728 SSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEK---AEDSSVEQLSTEIEPSRE 784
Db       1756 LESPNVATSSDSG---TRKSFMSLTTPTQSEREMTDSTPVFTETNTLENLGAQ---TTE 1808

Qy        785 ESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINED--------DPNVPAVVPLS--- 833
Db       1809 HSSI------------HQPGVQEGLTTLPRSPASVFMEQGSGEAAADPETTTVSSFSLNV 1856

Qy        834 EYNAVEREK---------MEAQGNTE-----------------LPQLLGAVYVSEELSK- 866
Db       1857 EY-AIQAEKEVAGTLSPHVETTFSTEPTGLVLSTVMDRVVAENITQTSREIVISERLGEP 1915

Qy        867 --------------------------TLVHTVS---VAVIDGT-RAVTSVEERSPSWIS 895
Db       1916 NYGAEIRGFSTGFPLEEDFSGDFREYSTVSHPIAKEETVMMEGSGDAAFRDTQTSPSTVP 1975

Qy        896 ASVTEPLEHTAGEAMP----------PVEEVTE------KDIIAEETPVLTQTLPEGKDA 939
Db       1976 TSV--HISHISDSEGPSSTMVSTSAFPWEEFTSSAEGSGEQLVTVSSSVV-PVLPSAVQK 2032

Qy        940 HDDMVTSEVD-FTSEAVTATE---TSEALRTEEV--TEASGAEETTDMVSAVSQLTDSPD 993
Db       2033 FSGTASSIIDEGLGEVGTVNEIDRRSTILPTAEVEGTKAPVEKEEVKVSGTVS--TNFPQ 2090

Qy        994 TTE--------EATPV-QEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRT 1044
Db       2091 TIEPAKLWSRQEVNPVRQEIESET--TSEEQIQEEKSFESPQNSPATEQTIFDSQTFTET 2148

Qy       1045 GSK-------------------------------------------ALEK------ 1051
Db       2149 ELKTTDYSVLTTKKTYSDDKEMKEEDTSLVNMSTPDPDANGLESYTTLPEATEKSHFFLA 2208
```

```
Qy   1052 ---------VEEVEEDSEVLASEKEKDVMPKG---PVQEAGAEHLAQGSETGQAT----- 1094
Db   2209 TALVTESIPAEHVVTDSPIKKEESTKH-FPKGMRPTIQESDTELLFSGLGSGEEVLPTLP 2267

Qy   1095 PESLEVPEV----------TADVDHVATCQVIKLQQL--MEQAVAPESSET--------- 1133
Db   2268 TESVNFTEVEQINNTLYPHTSQVESTSSDKIEDFNRMENVAKEVGPLVSQTDIFEGSGSV 2327

Qy   1134 --------LTDSETNGST--PLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEE--- 1180
Db   2328 TSTTLIEILSDTGAEGPTVAPLPFS-TDIGHPQNQTVRWAEEIQTS--RPQTITEQDSNK 2384

Qy   1181 -AATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQE--GEV 1237
Db   2385 NSSTAEINETTTSSTDFLAR-AYG-------FEMAKEFVTSAPKPSDLYYEPSGEGSGEV 2436

Qy   1238 DWLDGEKV--------KEEQEVFVHSG-----PNSQKAADVTYD-----SEVMGVAGCQE 1279
Db   2437 DIVDSFHTSATTQATRQESSTTFVSDGSLEKHPEVPSAKAVTADGFPTVSVMLPLHSEQN 2496

Qy   1280 KESTEVQSL--SLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDM 1337
Db   2497 KSSPDPTSTLSNTVSYERSTDGSFQDR-------FREFEDSTLKPNRKKPTENIIIDLDK 2549

Qy   1338 PSSERGKALGSLGGSPSLPD--QDKAGCIEV------------QVQSLDTTV-------- 1375
Db   2550 EDKDLILTITESTILEILPELTSDKNTIIDIDHTKPVYEDILGMQTDIDTEVPSEPHDSN 2609

Qy   1376 -------TQTAEAVEKVIETVVISETGE-SPECVGAHLLPAEKSSAT----------GGH 1417
Db   2610 DESNDDSTQVQEIYEAAVNLSLTEETFEGSADVLASYTQATHDESMTYEDRSQLDHMGFH 2669

Qy   1418 WTLQHAEDTVPLGPESQAE---------SIPIIVTPAPESTLHPDLQG------------ 1456
Db   2670 FT-----TGIP-APSTETELDVLLPTATSLPI---PRKSATVIPEIEGIKAEAKALDDMF 2720

Qy   1457 ---------------EI--SASQRERSEE--EDKPDAGPDADGKESTAIEKVL------- 1490
Db   2721 ESSTLSDGQAIADQSEIIPTLGQFERTQEEYEDKKHAGPSFQPEFSSGAEEALVDHTPYL 2780

Qy   1491 ---------KAEPEILELESKSNKIVLNVIQTAVDQFAR----TETAPETHAYDSQ---- 1533
Db   2781 SIATTHLMDQSVTEVPDVMEGSNPPYYTDTTLAVSTFAKLSSQTPSSPLTIYSGSEASGH 2840

Qy   1534 TQVP 1537
Db   2841 TEIP 2844
```

```
RESULT 13
NEST_RAT
ID   NEST_RAT       STANDARD;      PRT;  1805 AA.
AC   P21263;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Nestin.
GN   NES.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=90150286; PubMed=1689217;
RA   Lendahl U., Zimmerman L.B., McKay R.D.G.;
RT   "CNS stem cells express a new class of intermediate filament
RT   protein.";
RL   Cell 60:585-595(1990).
CC   -!- TISSUE SPECIFICITY: CNS STEM CELLS.
CC   -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
CC       IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M34384; AAA41685.1; -.
DR   PIR; A34736; A34736.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone.
FT   DOMAIN        1      7       HEAD.
FT   DOMAIN        8    314       ROD.
FT   DOMAIN      315   1805       TAIL.
FT   DOMAIN        8     43       COIL 1A.
FT   DOMAIN       44     55       LINKER 1.
FT   DOMAIN       56    151       COIL 1B.
FT   DOMAIN      152    174       LINKER 12.
FT   DOMAIN      175    193       COIL 2A.
FT   DOMAIN      194    196       LINKER 2.
FT   DOMAIN      197    314       COIL 2B.
SQ   SEQUENCE   1805 AA;  198744 MW;  B40EE14717E0998D CRC64;

  Query Match             4.8%;  Score 388;  DB 1;  Length 1805;
  Best Local Similarity   21.3%;  Pred. No. 1.3e-06;
  Matches  319;  Conservative  233;  Mismatches  543;  Indels  402;  Gaps   75;

Qy     13 EQPAGSDT---PSELV-----LSGHGPAAEASGAAGDPADADPATKLPQKNGQLSSVNGV 64
Db    440 EEPGGKQQGHFPDDLTSLATNLNPHHPTLEAKD--GESSESRVSSIFQEDEGQIWEL--V 495

Qy     65 AEQGDVHVQEENQEGQ--EEEVVDEDVGQRESEDVREKDRVEEMAAN---STAVEDITKD 119
Db    496 EKEADIEVKVENSSAQKTQESGLDTEETQ-DSQGPLQKETLKALGEEPLMSLKIQNYETA 554

Qy    120 GQE------------ETSEIIEQIP---ASENNVEEMVQPAESQANDVGFKKVFKFVGF 163
Db    555 GKENCNSSTEGHLGTLEGPEKEKQIPLKSLEEKNVE------SEKTLENGVPVLSELLGK 608

Qy    164 KFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVET----------AVGESA 213
Db    609 EDTRTEDQELMSPKG---TLKRFSSLGKESQEVVRPSKEGNLESWTAFKEESQHPLGFPG 665

Qy    214 SKESELKQSTEKQEGTL----KQEQSSTEIPLQAESDQ-AAEEEAKDEGE------EKQE 262
Db    666 AEDQMLERLVEKEDQSFPRSPEEEDQEACRPLQKENQEPLGYEEA--EGQILERLIEKES 723

Qy    263 KEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEK 322
Db    724 QESLRSPEEE----DQEAGRSLQK-------ENQEPLGYEEAEDQMLERLIEKESQESLK 772

Qy    323 VDEEEKEKTEPASEEQE--------------PAEDTDQARLSA-----DYEKVELPLE-- 361
Db    773 SPEENQRIGKPLERENQKSLRYLEENQETFVPLESRNQRPLRSLEVEEEEQRIVKPLEKV 832

Qy    362 --DQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGEAEGGV 419
Db    833 SQDSLGSL---AEENVQPLR---YLEEDDCINKSLLE------DKTHKSLGSLEDRNGDS 880

Qy    420 VV----EGTGESL-PPEK---------LAEPQEVPQEAEPAEELMKSREMCVSGGDHTQL 465
Db    881 IIIPQESETQVSLRPPEEEDQRIVNHLEKESQEFSRSSEEEEQVM---ERSLEGENHESL 937

Qy    466 TDLSPE----EKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKG 521
Db    938 SSVEKEDQMVESQLEKESQDSGKSLED-ESQETF----GPLEK-ENAESLRSLAGQDQ-- 989

Qy    522 KRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQD-GEAE 580
```

```
Db     990 -----------EEQKLEQETQQTLRAVGNEQMAVSPP--------EKVDPELPKPLGNDQ 1030

Qy     581 EGTTSDGEKKREGITPWASFKKMVTPKK---RVRRPSESDKEEELEKVKS----ATLSST 633
           |   | |:: :| :         | : | |        :  | |:  ||:||: ||         | ||
Db    1031 EIARSLGKENQESLVSLKE-KGIETVKSLETEIIEPLET-AEEDLERRKSIDTQEPLWST 1088

Qy     634 D---STVSEMQDE----VKTVGEEQKP----EEPKRRVDTSVSWEALICVGSSKKRARKA 682
           :      ||    :||      : :| | ::     |:   : : :    |      |  |:: :
Db    1089 EVARETVEPPEDEPPGSLGSVDENRETLTSLEKESQELSSLGKWNVETRVEDSQQCLQVE 1148

Qy     683 SSSDDEGGPRTL---------GGDSHRAEEASKDKEAGTDAVPAST-------------- 719
              :|     :|              |:   |  |:   :  |  :|   |:|
Db    1149 EGLQEEQHQESLREVKQELPSSGNQQRWEDVVEGKAVGQEAPLATTGVGTEDKAELHLRG 1208

Qy     720 ---QEQDQAQ--------------GSSSPEPAGSPSEG----EGVSTWESFKRLVTPRKK 758
               :|:  |:                  ||| |:     |:|     || :        |  |   :
Db    1209 QGGEEEAAAEGELLQDIVGEAWSLGSSEPKEQRVPAEALDNLEGGA-------LEVPVAQ 1261

Qy     759 SKSKLEEKAED--SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGP 816
           |  :: |: ||     : || | |:      |  :     ::                  :||   :|| |
Db    1262 SMPEVTERDEDRAQAGEQDSIEVTLGLEAARTGLEL-------------EQEVVGLED--P 1307

Qy     817 VEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVS-EELSKTLVHTVSVA 875
              ::      |:     | | :      |:   :  | :|     :|  |:   ||:      || ||       :
Db    1308 RHFAREEAIPPS---LGEESVKAKIAQGLEGPGKEPKEAGALDSGILELPKTSSEALECQ 1364

Qy     876 VIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEET--------- 926
             : : ::          ||    |              |   :|      :|   |      ||:|    |:
Db    1365 GHEESESMEGWEEEEAS------LETSDHEGSDAPQPRPPETEEDEGAQAALTAPGPKLL 1418

Qy     927 ------PVLTQTLPEGKDAHDDMVTSE--------VDFTSEAVTATETSEALRTEEVTEA 972
                   |:||           |||:     :|                :    |||:    |         |: |
Db    1419 EPCSPIPILT-------DAHELQPQAEGIQEAGWQPEAGSEALERVENEPEFGLGEIPEG 1471

Qy     973 -----SGAEETTDMVSAVSQLTDSPDTTEEATP----VQEVESGVLDTEEEERQTQAILQ 1023
                  | ||          |:   |  :|   ::||      ::     |      |     |:|         |
Db    1472 LQDWEEGREE--------SEADDLGETLPDSTPLGLYLRSPASPKWDLAGEQR-----LS 1518

Qy    1024 AVADKVKEE--SQVPATQTVQRTGSKALEKVEEVEEDSEVLASEK------EKDVMPKGP 1075
               |  ||:      ||| |        |           :  || :        |:||:         |      ::|    |
Db    1519 PQGDAGKEDWGPAVPAAQ-----GLSGPPEEEEEQGHGSDLSSEEFEDLGTEASLLPGVP 1573

Qy    1076 VQEAGAEHLAQ-------------GSETGQATPESL--EVPEVTADVDHVAT-------- 1112
              :       |:|: |                                  |    |   |    |    |    :|| :       :
Db    1574 KEV--ADHVGQVPPVLQPACWDQGGESDGFADEEESGEEGEEEDADEEGAESGAQWWGSG 1631

Qy    1113 -----CQVIKLQQLMEQAVAPESSETLTDSETNGST------PLADSDTADGTQQDETID 1161
                   |:|    :  |    :                 |  |        |:                :|:|:  :  :|:
Db    1632 ASGGGCKVQDIAQRGDPVQESVGVSGLWDDGLRGAAANVPALEMVSQDSAEPSGSEES-- 1689

Qy    1162 SQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAA 1221
                     |:       :   |||:         |  :|     |   :|           |     :     |:   |      |
Db    1690 ---ESASLEGEEGQVTDH--LDAPQEVTSMVPGVGDAFDIGGQSPNLD------------ 1732

Qy    1222 AVPVLAKTEVG---QEGEVDWLDGEKVKEEQEVF-----------------VHSGPN 1258
           :   |   |  |  |         ||:|     |||:      |:||:                        || ||:
Db    1733 SEQVNGKMENGLEQAEGQV-VLDGD---EDQELLLQGQEVGALKVPLVASPVHLGPS 1785
```

```
RESULT 14
MAP2_RAT
ID   MAP2_RAT       STANDARD;      PRT;  1861 AA.
AC   P15146;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Microtubule-associated protein 2 (MAP 2) (MAP2B) [Contains: MAP2C].
GN   MAP2.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC   STRAIN=Wistar; TISSUE=Brain;
RX   MEDLINE=90251471; PubMed=2339070;
RA   Kindler S., Schwanke B., Schulz B., Garner C.C.;
RT   "Complete cDNA sequence encoding rat high and low molecular weight
RT   MAP2.";
RL   Nucleic Acids Res. 18:2822-2822(1990).
RN   [2]
RP   SEQUENCE OF 1-1694 AND 1726-1861 FROM N.A.
RC   STRAIN=Wistar; TISSUE=Brain;
RX   MEDLINE=91060576; PubMed=2174050;
RA   Kindler S., Schulz B., Goedert M., Garner C.C.;
RT   "Molecular structure of microtubule-associated protein 2b and 2c from
RT   rat brain.";
RL   J. Biol. Chem. 265:19679-19684(1990).
RN   [3]
RP   SEQUENCE OF 1-151; 1515-1694 AND 1726-1861 FROM N.A.
RX   MEDLINE=90221819; PubMed=2326166;
RA   Doll T., Papandrikopoulou A., Matus A.;
RT   "Nucleotide and amino acid sequences of embryonic rat MAP2c.";
RL   Nucleic Acids Res. 18:361-361(1990).
RN   [4]
RP   DISCUSSION OF SEQUENCE.
RX   MEDLINE=89365159; PubMed=2770869;
RA   Papandrikopoulou A., Doll T., Tucker R.P., Garner C.C., Matus A.;
RT   "Embryonic MAP2 lacks the cross-linking sidearm sequences and
RT   dendritic targeting signal of adult MAP2.";
RL   Nature 340:650-652(1989).
RN   [5]
RP   SEQUENCE OF 1695-1725 FROM N.A.
RX   MEDLINE=94110302; PubMed=8282767;
RA   Doll T., Meichsner M., Riederer B.M., Honegger P., Matus A.;
RT   "An isoform of microtubule-associated protein 2 (MAP2) containing
RT   four repeats of the tubulin-binding motif.";
RL   J. Cell Sci. 106:633-640(1993).
CC   -!- FUNCTION: THE EXACT FUNCTION OF MAP2 IS UNKNOWN BUT MAPS MAY
CC       STABILIZE THE MICROTUBULES AGAINST DEPOLYMERIZATION. THEY ALSO
CC       SEEM TO HAVE A STIFFENING EFFECT ON MICROTUBULES.
CC   -!- ALTERNATIVE PRODUCTS: VARIOUS FORMS OF MAP2 ARE PRODUCED BY
CC       ALTERNATIVE SPLICING OF THE SAME GENE. MAP2C, THE LOW MOLECULAR
CC       FORM OF MAP2, LACKS THE CENTRAL DOMAIN OF MAP2A/B.
CC   -!- DEVELOPMENTAL STAGE: MAP2C IS EXPRESSED DURING EMBRYONIC BRAIN
CC       DEVELOPMENT AND UNTIL POSTANATAL DAY 10. MAP2B IS EXPRESSED
CC       THROUGHOUT BRAIN DEVELOPMENT.
CC   -!- SIMILARITY: CONTAINS 3 OR 4 TAU/MAP REPEATS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X51842; CAA36135.1; -.
DR   EMBL; X17682; CAA35667.1; -.
DR   EMBL; X71487; CAA50588.1; -.
DR   PIR; S07887; S07887.
DR   PIR; S10003; S10003.
DR   PIR; A37981; A37981.
DR   InterPro; IPR001084; Tubulin_Tau.
DR   Pfam; PF00418; tubulin-binding; 4.
DR   PROSITE; PS00229; TAU_MAP; 3.
KW   Microtubules; Repeat; Alternative splicing; Calmodulin-binding.
FT   DOMAIN      1454   1474       CALMODULIN-BINDING (POTENTIAL).
FT   REPEAT      1664   1694       TAU/MAP MOTIF.
FT   REPEAT      1695   1725       TAU/MAP MOTIF.
FT   REPEAT      1726   1756       TAU/MAP MOTIF.
FT   REPEAT      1757   1788       TAU/MAP MOTIF.
FT   VARSPLIC     152   1514       MISSING (IN ISOFORM MAP2C).
FT   VARSPLIC    1695   1725       MISSING (IN ISOFORM WITH 3 TAU/MAP
FT                                 REPEATS).
SQ   SEQUENCE   1861 AA;  202409 MW;  42DCF116D21EF54E CRC64;
```

```
Query Match             4.8%;  Score 388;  DB 1;  Length 1861;
Best Local Similarity   19.8%;  Pred. No. 1.3e-06;
Matches  306;  Conservative  220;  Mismatches  590;  Indels  432;  Gaps   66;

Qy        7 TEQRSPEQPAGSDTPSELVL---SGHGPAAEASGAAGDPADADPATKLPQKNGQLSSVNG 63
              ||   |   |     :   ::  |     ||    :  |:|        | :|       |    :
Db      335 TEGSQPFPAPVFFQSDDKMSLQDTSGSATSKESS------KDEEPQKDKADKVADVPVSEA 388

Qy       64 VAEQGDVHVQE------ENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDIT 117
                 ||||           ||   |:|:     |::   :   :  |:|             |   |   |
Db      389 TTVLGDVHSPAVEGFVGENISGEEKGTTDQEKKETSTPSVQE--------PTLTETEPQT 440

Qy      118 KDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDT 177
              |    ||||::              ::||  |     |                     | |::|:
Db      441 K--LEETSKV---------SIEETVAKEEESL-------------------KLKDDKAGV 470

Qy      178 VQLLT---VKKDEGEGAEASVGA--GDHQEPSVETAVGESASKESELKQSTEKQEGTLKQ 232
              :|  |      |:: :|  | :: |    |      |:|  || ::|      |:: |  :  |      ::
Db      471 IQTSTEQSFSKEDQKGQEQTIEALKQDSFPISLEQAVTDAAMATKTLEKVTSEPEAVSEK 530

Qy      233 EQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSP--ESPSSPVNSETTSSFKKFFTH 290
              :          :|    ::   :::| ||                        ||: |:  :  |
Db      531 RE------IQGLFEEDIADKSKLEGAGSATVAEVEMPFYEDKSGMSKYFETSALKEDVTR 584

Qy      291 G---WAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQA 347
                : : : :   : : :: |:|    : :|          :|||      ||:    ||:    :
Db      585 STGLGSDYYELSDSRGNAQESLDTVSPKNQQ-------DEKELLAKASQPSPPAHEAGYS 637

Qy      348 RLSADY---EKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEK 404
               |:  |          ||| |           :| :|:    :   :|:  ||  :  |  :     :: :|
Db      638 TLAQSYTSDHPSELPEE------PSSPQERMFTIDPKVYGEKRDLHSKNKDDLTLS---- 687

Qy      405 TEEEQGGGGEAEGGVVVE---------------------GTGESLPPEKLAEPQEVPQE 442
                  | ||  :         :|                              |    |    |    |
Db      688 --RSLGLGGRS----AIEQRSMSINLPMSCLDSIALGFNFGRGHDLSP----------- 729

Qy      443 AEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPL 502
                 | ::: :      :   ||         :|   :  :|      |        :    |:   ||  |
Db      730 --LASDILTNTSGSMDEGD-DYLPPTTPAVEKIPCFP---IESKEEEDKTEQAKVTGGQT 783

Qy      503 KKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPE- 561
              :: :||                                  |||   |   |        :   :|:  ||
Db      784 TQVETSS-----------------------------ESPFPAKEYYKNGTVMAPDLPEM 813

Qy      562 -----------------ETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMV 604
                                  |    :   |  ||       ||    |               |:  |  |      :  |
Db      814 LDLAGTRSRLASVSADAEVARRKSVPSEAV---VAESST---------GLPPVADDSQPV 861

Qy      605 TPKKRVRRPSESDKEEELEKVKSATLSS-TDSTVSEMQDEVKTVGE-----EQKPEEPKR 658
              |               : :|| :       :    |      |  :||        ||       |       :: :|
Db      862 KP-----------DSQLEDMGYCVFNKYTVPLPSPVQDSENLSGESGSFYEGTDDKVRR 909

Qy      659 RVDTSVSWEALICVG-SSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAG--TDAV 715
              :  | :|     ||| |    ::   |:  :  :::   |    |  :     |     |    |   :|::|       | |
Db      910 DLATDLS---LIEVKLAAAGRVKDEFTAEKEASPPSSADKSGLSREFDQDRKANDKLDTV 966

Qy      716 PASTQEQ-DQAQGSSSPEPAGSPSE--GEGVSTWE--SFKRLVTPRKKSKSKLEEKAED- 769
              ::|    |    :   :     |      |    | | |  |:|   |    |  |  :|      :|     |:||
Db      967 LEKSEEHVDSKEHAKESEEVGDKVELFGLGV-TYEQTSAKELIT----TKETAPERAEKG 1021

Qy      770 -SSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDD----- 823
               ||| ::: |:| :                       |      |       |       :|    |:|     |
Db     1022 LSSVPEVA-EVETT-----------------TKADQGLDVAAKKDDQSPLDIKVSDFGQMA 1064

Qy      824 --PNVPAVVPLS-EYNAVEREKMEAQGNTELPQLLG--------------AVYVSEELSK- 866
                 :|  |    :  ::     ::   : ::       |        :|                      |  |:::|
Db     1065 SGMSVDAGKTIELKFEVDQQLTLSSEAPQETDSFMGIESSHVKDGAKVSETEVKEKVAKP 1124

Qy      867 TLVHTVSV--------------AVIDGTRAVTSVEERSPSW--------ISASVTEPLEH 904
                |||    :|                          ::     :          :|  ||                :    ||   |
Db     1125 DLVHQEAVDKEESYESSGEHESLTMESLKPDEGKKETSPETSLIQDEVALKLSVEIPCPP 1184

Qy      905 TAGEAMPPVEEVTEKDI----IAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
                 ||     ::|   |    :          :  :|          |    :|                :  |:|             :|
Db     1185 PVSEADSSIDEKAEVQMEFIQLPKEESTETPDIP--------AIPSDVTQPQPEAVVSEP 1236

Qy      961 SEALRTEEVTEASGAEETTDMVSAVSQLTD--SPDTTEEATPVQEVESGVLDTEEEERQT 1018
               :|       ||   ||   |      :             |          |:||     |  :  ||                  |       :
Db     1237 AEVRGEEEEIEAEGEYDKLLFRSDTLQITDLLVPGSREEF-----------VETCPGEH 1284

Qy     1019 QAILQAVADKVKEESQVPATQTVQRTGSKALEKVE-----EVEEDSEVLASEKEKDVMPK 1073
               :  ::::|         |:    :          ||         |                   :   ||:            ::|   :|:
Db     1285 KGVVESVV--TIEDDFITVVQTTTDEGELGSHSVRFAAPVQPEEERRPYPHDEELEVL-- 1340

Qy     1074 GPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSET 1133
                   ||  ||        ||           ||||   ||                                     :      :||
Db     1341 -MAAEAQAEP-KDGSPDAPATPEKEEVP----------------------FSEYKTET 1374

Qy     1134 LTD--SETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPST 1191
                 |      ||           :    |:|:                 :|:||                :      |:     |     |||   :
Db     1375 YDDYKDETTIDDSIMDADSL-------WVDTQDD-------DRSILTEQLETIPKEERAE 1420

Qy     1192 LPNNVPAQEEHGEE----PGRD-VLEPTQQELTAAVPVLAKTEVGQEGEVDWLDGEKVK 1246
                             |:  |:|  :|            ||    :    |  ::|:           :::   ||  ::     |            ::
Db     1421 KEARRPSLEKHRKEKPFKTGRGRISTPERREVAKKEPSTVSRDEVRRKKAV--YKKAELA 1478

Qy     1247 EEQEVFVHS----------------------------------GPNSQKAADVTYDSEV 1271
               :|   ||     ||                                                         |::  |  |             |
Db     1479 KESEVQAHSPSRKLILKPAIKYTRPTHLSCVKRKTTATSGESAQAPSAFKQAK---DKVT 1535

Qy     1272 MGVAGCQEKESTEVQSLSL-----------EEG--EMETDVEKEKRETKPEQVSEEGEQE 1318
                |:        ||  |:   :     |:                           ||            :   :  :        :|   |:  |:   :        |:
Db     1536 DGITKSPEKRSSLPRPSSILPPRRGVSGDREENSFSLNSSISSARRTTRSEPIRRAGKSG 1595

Qy     1319 TAAPEHEGTYGKPVLTLDMPSSERGKALGSLG--GSPSLPDQDKAGCI 1364
                |:  |          |  |        :|      |    |         :     |:   |           |      |        |:|  :
Db     1596 TSTPT---TPGSTAITPGTPPSYSSRTPGTPGTPSYPRTPGTPKSGIL 1640
```

```
RESULT 15
NFH_MOUSE
ID   NFH_MOUSE      STANDARD;      PRT;  1087 AA.
AC   P19246; Q61959;
DT   01-NOV-1990 (Rel. 16, Created)
DT   01-AUG-1992 (Rel. 23, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Neurofilament triplet H protein (200 kDa neurofilament protein)
DE   (Neurofilament heavy polypeptide) (NF-H).
GN   NEFH OR NFH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89121513; PubMed=3220257;
RA   Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
RA   Mushynski W.;
RT   "Sequence and structure of the mouse gene coding for the largest
RT   neurofilament subunit.";
RL   Gene 68:307-314(1988).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=89089138; PubMed=3145094;
RA   Shneidman P.S., Carden M.J., Lees J.F., Lazzarini R.A.;
RT   "The structure of the largest murine neurofilament protein (NF-H) as
RT   revealed by cDNA and genomic sequences.";
RL   Brain Res. 464:217-231(1988).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Swiss Webster; TISSUE=Brain;
RA   Carden M.J.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC       AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
```

```
CC       NF-H HAS AN IMPORTANT FUNCTION IN MATURE AXONS THAT IS NOT
CC       SUBSERVED BY THE TWO SMALLER NF PROTEINS.
CC   -!- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPEPTIDE K-S-P, NFH IS
CC       PHOSPHORYLATED ON A NUMBER OF THE SERINES IN THIS MOTIF. IT IS
CC       THOUGHT THAT PHOSPHORYLATION OF NFH RESULTS IN THE FORMATION OF
CC       INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC       OF AXONAL CALIBER.
CC   -!- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC       OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC       LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC       COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC   -!- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN IN POSITIONS 534
CC       TO 716 AND IS SHORTER DUE TO FRAMESHIFTS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between  the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC   the European Bioinformatics Institute.  There are no  restrictions on  its
CC   use  by  non-profit  institutions as long  as its content  is  in  no  way
CC   modified and this statement is not removed.  Usage  by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M24496; AAA39813.1; -.
DR   EMBL; M23349; AAA39813.1; JOINED.
DR   EMBL; M24494; AAA39813.1; JOINED.
DR   EMBL; M24495; AAA39813.1; JOINED.
DR   EMBL; M35131; AAA39809.1; ALT_FRAME.
DR   EMBL; Z31012; CAA83229.1; -.
DR   PIR; JT0368; QFMSH.
DR   PIR; A43778; A43778.
DR   MGD; MGI:97309; Nfh.
DR   InterPro; IPR001664; IF.
DR   Pfam; PF00038; filament; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Repeat.
FT   DOMAIN        1     97       HEAD.
FT   DOMAIN       98    408       ROD.
FT   DOMAIN      409   1087       TAIL.
FT   DOMAIN      436    517       GLU-RICH (ACIDIC).
FT   DOMAIN      519    886       50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
FT   DOMAIN      887   1087       GLU/LYS-RICH.
FT   DOMAIN       98    129       COIL 1A.
FT   DOMAIN      130    141       LINKER 1.
FT   DOMAIN      142    239       COIL 1B.
FT   DOMAIN      240    261       LINKER 12.
FT   DOMAIN      262    283       COIL 2A.
FT   DOMAIN      284    287       LINKER 2.
FT   DOMAIN      288    408       COIL 2B.
FT   CONFLICT    133    133       K -> QA (IN REF. 2 AND 3).
FT   CONFLICT    199    199       A -> AR (IN REF. 2 AND 3).
FT   CONFLICT    281    281       S -> T (IN REF. 2 AND 3).
FT   CONFLICT    492    492       L -> G (IN REF. 2 AND 3).
FT   CONFLICT    551    551       P -> PREAKSP (IN REF. 3).
FT   CONFLICT    689    712       MISSING (IN REF. 3).
FT   CONFLICT    714    714       G -> A (IN REF. 3).
FT   CONFLICT    814    814       V -> M (IN REF. 2 AND 3).
FT   CONFLICT    843    843       T -> N (IN REF. 2 AND 3).
SQ   SEQUENCE   1087 AA;  116612 MW;  57BAC76A38ED1CB9 CRC64;

  Query Match             4.8%;  Score 385.5;  DB 1;  Length 1087;
  Best Local Similarity   23.8%;  Pred. No. 9.3e-07;
  Matches  219;  Conservative  139;  Mismatches  374;  Indels  189;  Gaps   43;

Qy         26 LSGHGPA-AEASGAAGDPADADPATKLPQKNGQLSSVNGVAEQGDVHVQE----ENQEGQ 80
              : | | | |:|   | |       |  : |  :   ||     | | |   :  :|
Db        240 IQGCGAAQAQAQAEARDALKCDVTSALREIRAQL---EGHAVQSSLQSEEWFRVRLDRLS 296

Qy         81 EEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDI--TKDGQE-ETSEIIEQIPASENN 137
              |     |: |   :    |::  |   |   :: | :| :|  :  ||:   | : ||: :: |    :
Db        297 EAAKVNTDAMRSAQEEITEYRR--QLQARTTELEALKSTKESLERQRSELEDRHQADIAS 354

Qy        138 VEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVK---------KDE 187
              : : :|   :|:  :                |:  :        |     |     || ||                   :
Db        355 YQDAIQQLDSELRNT----------KWEMAAQLREYQD---LLNVKMALDIEIAAYRKL 400

Qy        188 GEGAEASVGAGDH---------QEPSVETAVGESASKESELKQSTEKQ----EGTLKQEQ 234
              || |   :| |                  : ||: | :    :  :  :: : :||:     ||   :: :
Db        401 LEGEECRIGFGPSPFSLTEGLPKIPSISTHIKVKSEEMIKVVEKSEKETVIVEGQTEEIR 460

Qy        235 SSTEIPLQAESD-QAAEEEAKDEGEEKQEKE---PTKSPESPSSPVNSETTSSFKKFFTH 290
              :   :  : : :  |   |  |   :||||:|:|      |    |      ::       || |   |:
Db        461 VTEGVTEEEDKEAQGQEGEEAEEGEEKEEEELAAATSPPAEEAASPEKETKSRVKE---- 516

Qy        291 GWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDE-----EEKEKTEPASEEQEPAEDTD 345
                    |          ||   :       ||  :      ||:      |        | |     | |     |||
Db        517 ----EAKSPGEAKSPGEAKSPAEAKSPGEAKSPGEAKSPGEAKSPAEPKS----PAEPKS 568

Qy        346 QARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKT 405
              |    :    |                  |: ::  ||   |         |      :   |    |   :   ||         |:
Db        569 PAEAKSPAEPKSPATVKSPGEAKSPSEAKSPAEAKSPAEAKSPAEAKSPAEA------KS 622

Qy        406 EEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAE---PAEELMKSREMCVSGGDH 462
                    ||:        |:   ||:   |  :    |  |     ||     |||     ||       |  |:
Db        623 PAEAKSPAEAKSPATVKSPGEAKSPSEAKSPAEAKSPAEAKSPAE--AKSPAEVKSPGEA 680

Qy        463 TQLTDLSPEEKTLPKHPEGIVSEVEMLSSQE-----RIKVQG---SPLKKLFSSSGLKKL 514
                    || |     || |      |   |: |   |          :|    ||     : |  :   |
Db        681 K-----SPAE---PKSPAEAKSPAEVKSPAEAKSPAEVKSPGEAKSP-AAVKSPAEAKSP 731

Qy        515 SGKKQKGKRGGGD-------EEPGEYQH-IHTESPESADE--QKGESSASSPEEPEETT 564
              :  |   |:        |:          :  |  |  :    |   :|||  |      ::|    |   :   :  ||:
Db        732 AAVKSPGEAKSPGEAKSPAEAKSPAEAKSPIEVKSPEKAKTPVKEGAKSPAEAKSPEKAK 791

Qy        565 CLEKGPLEAPQDGEAEEGTTS---DGEKKREGITPW--------------ASFKKMVTPK 607
                |   ::  |   : ::  |        |       :|   |   |                      |:     :  |
Db        792 SPVKEDIKPPAEAKSPEKAKSPVKEGAKPPEKAKPLDVKSPEAQTPVQEEATVPTDIRPP 851

Qy        608 KRVRRPS-ESDKEEELEKVKSATLSSTDSTVSEMQDEVKT-VGEEQKPEEPKRRVDTSVS 665
              ::|:  |:  |   |   |  |  |:  |       |:   ::|||:  |  ||  |  :||  ::|:
Db        852 EQVKSPAKEKAKSPEKEEAK------TSEKVAPKKEEVKSPVKEEVKAKEPPKKVEEE-- 903

Qy        666 WEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQA 725
                                               :||          |:|:  ||  ||   :|        :|:  :
Db        904 --------------------------KTLPTPKTEAKESKKD-EAPKEAPKPKVEEKKE- 935

Qy        726 QGSSSPEPAGSPSEGEGVSTWESFKRLV----TPRK---KSKSKLEEKAEDSSVEQLSTE 778
                : : :|   |:|  :         |  |  :          ||  |      |::|  :||  |  :    |    |:
Db        936 --TPTEKPKDSTAEAKKEEAGEKKKAVASEEETPAKLGVKEEAKPKEKTETTKTEAEDTK 993

Qy        779 I-EPSR-EESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVE-------INEDDPNVPAV 829
                 |||:   |:        |: :|      :|:  |  |:|:  |  |    | |                  :  |||   ::
Db        994 AKEPSKPTETEKPKKEEMPAAPEKK-DTKEEK-TTESRKPEEKPKMEAKVKEDDKSLS-- 1049

Qy        830 VPLSEYNAVEREKMEAQGNTE 850
                   |  :   :  ||  |      :|:
Db       1050 ---KEPSKPKTEKAEKSSSTD 1067


Search completed: December 13, 2002, 00:13:37
Job time : 85 secs
```

```
OM protein - protein search, using sw model

Run on:          December 13, 2002, 00:00:30 ; Search time 94 Seconds
                                          (without alignments)
                                          3498.417 Million cell updates/sec

Title:           US-09-902-432-4
Perfect score:   8073
Sequence:        1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters:      671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_21:*
                 1:  sp_archea:*
                 2:  sp_bacteria:*
                 3:  sp_fungi:*
                 4:  sp_human:*
                 5:  sp_invertebrate:*
                 6:  sp_mammal:*
                 7:  sp_mhc:*
                 8:  sp_organelle:*
                 9:  sp_phage:*
                 10:  sp_plant:*
                 11:  sp_rodent:*
                 12:  sp_virus:*
                 13:  sp_vertebrate:*
                 14:  sp_unclassified:*
                 15:  sp_rvirus:*
                 16:  sp_bacteriap:*
                 17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	8073	100.0	1596	11	Q62766	Q62766 rattus norv
2	8030	99.5	1596	11	Q9Z1F7	Q9z1f7 rattus norv
3	6512.5	80.7	1684	11	Q9WTQ5	Q9wtq5 mus musculu
4	6035.5	74.8	1579	11	Q99MP1	Q99mp1 mus musculu
5	1805	22.4	2471	13	Q9DDN8	Q9ddn8 xenopus lae
6	642	8.0	2768	5	Q9VC00	Q9vc00 drosophila
7	635	7.9	5458	5	Q9U459	Q9u459 plasmodium
8	588	7.3	17352	5	Q95YM2	Q95ym2 procambarus
9	585.5	7.3	1829	5	Q22248	Q22248 caenorhabdi
10	574	7.1	5327	5	O76891	O76891 drosophila
11	564.5	7.0	2910	10	Q9FND5	Q9fnd5 arabidopsis
12	558	6.9	16215	5	Q9NFS3	Q9nfs3 drosophila
13	557.5	6.9	6815	5	Q9I7U4	Q9i7u4 drosophila
14	517	6.4	2081	10	Q9LH98	Q9lh98 arabidopsis
15	501.5	6.2	3484	5	P91257	P91257 caenorhabdi
16	493.5	6.1	2083	5	Q9N435	Q9n435 caenorhabdi
17	489.5	6.1	1852	3	Q9C2H4	Q9c2h4 neurospora
18	489	6.1	6632	5	O01761	O01761 caenorhabdi
19	489	6.1	6994	5	Q17343	Q17343 caenorhabdi
20	488	6.0	5170	5	Q17490	Q17490 caenorhabdi
21	485	6.0	2484	5	Q9U347	Q9u347 caenorhabdi
22	485	6.0	2607	5	Q23187	Q23187 caenorhabdi
23	485	6.0	3263	5	Q9I7U3	Q9i7u3 drosophila
24	485	6.0	6632	5	Q17362	Q17362 caenorhabdi
25	472	5.8	1514	5	Q8SY55	Q8sy55 drosophila
26	466.5	5.8	1871	10	Q9SRD5	Q9srd5 arabidopsis
27	463	5.7	1661	5	Q06166	Q06166 plasmodium
28	462	5.7	1786	5	Q9U0P0	Q9u0p0 plasmodium
29	456.5	5.7	1804	11	Q9Z1Q1	Q9z1q1 mesocricetu
30	452	5.6	3239	5	Q8T102	Q8t102 bombyx mori
31	442	5.5	1326	16	Q9L2C3	Q9l2c3 streptomyce
32	440.5	5.5	1812	5	Q9VZD9	Q9vzd9 drosophila
33	438	5.4	1128	5	Q26947	Q26947 trypanosoma
34	438	5.4	1616	5	Q9VSJ0	Q9vsj0 drosophila
35	434	5.4	1510	5	Q25920	Q25920 plasmodium
36	431	5.3	3257	5	Q9V736	Q9v736 drosophila
37	430	5.3	3111	5	Q9VH10	Q9vh10 drosophila
38	429	5.3	2748	5	Q9V4P4	Q9v4p4 drosophila
39	429	5.3	3261	4	Q9Y556	Q9y556 homo sapien
40	429	5.3	3664	4	Q96T58	Q96t58 homo sapien
41	425	5.3	1110	13	Q91255	Q91255 petromyzon
42	420.5	5.2	4900	5	Q9N541	Q9n541 caenorhabdi
43	419	5.2	3187	11	Q63714	Q63714 rattus norv
44	414	5.1	1422	6	Q95KU4	Q95ku4 canis famil
45	414	5.1	5085	11	Q9JKS6	Q9jks6 rattus norv

ALIGNMENTS

```
RESULT 1
Q62766
ID   Q62766      PRELIMINARY;      PRT;  1596 AA.
AC   Q62766;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   SSeCKS.
GN   322.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FISHER; TISSUE=EMBRYO;
RX   MEDLINE=95257957; PubMed=7739556;
RA   Lin X., Nelson P.J., Frankfort B., Tombler E., Johnson R.,
RA   Gelman I.H.;
RT   "Isolation and characterization of a novel mitogenic regulatory gene,
RT   322, which is transcriptionally suppressed in cells transformed by src
RT   and ras.";
RL   Mol. Cell. Biol. 15:2754-2762(1995).
DR   EMBL; U23146; AAA79517.1; -.
DR   InterPro; IPR001573; PkinA_anch.
SQ   SEQUENCE   1596 AA;  172498 MW;  CF78F2D161D9C7E2 CRC64;

  Query Match             100.0%;  Score 8073;  DB 11;  Length 1596;
  Best Local Similarity   100.0%;  Pred. No. 1.9e-314;
  Matches 1596;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60

Qy         61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
```

```
Qy     121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy     181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240

Qy     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300

Qy     301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360

Qy     361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420

Qy     421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480

Qy     481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540

Qy     541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600

Qy     601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Qy     661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720

Qy     721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780

Qy     781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840

Qy     841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Qy     901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Qy     961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Qy    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080

Qy    1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140

Qy    1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200

Qy    1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260

Qy    1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320

Qy    1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380

Qy    1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440

Qy    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500

Qy    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560

Qy    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
           ||||||||||||||||||||||||||||||||||||
Db    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596


RESULT 2
Q9Z1F7
ID   Q9Z1F7      PRELIMINARY;      PRT;  1596 AA.
AC   Q9Z1F7;
DT   01-MAY-1999 (TrEMBLrel. 10, Created)
DT   01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PKC binding protein and substrate.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Chapline C., Mousseau B., Ramsay K., Jaken S.;
RT   "Identification of a major protein kinase C binding protein and
RT   substrate in rat embryo fibroblasts: Decreased expression in
RT   transformed cells.";
RL   J. Biol. Chem. 0:0-0(1995).
DR   EMBL; U41453; AAD03788.1; -.
DR   InterPro; IPR001573; PkinA_anch.
SQ   SEQUENCE   1596 AA;  172412 MW;  E6C4F4074F1725D3 CRC64;

  Query Match             99.5%;  Score 8030;  DB 11;  Length 1596;
  Best Local Similarity   99.6%;  Pred. No. 9.9e-313;
  Matches 1589;  Conservative    3;  Mismatches    4;  Indels    0;  Gaps    0;

Qy       1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60

Qy      61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120

Qy     121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy     181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240

Qy     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300

Qy     301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPL 360

Qy     361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 EDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVV 420

Qy     421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     421 VEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPE 480

Qy     481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     481 GIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTE 540

Qy     541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     541 SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASF 600

Qy     601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     601 KKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRV 660

Qy     661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     661 DTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQ 720

Qy     721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     721 EQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIE 780

Qy     781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER 840
           ||||||||||||||||||||||||||||||||||||||||||||:||||||||||:||||
Db     781 PSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPDVPAVVPLSEYDAVER 840

Qy     841 EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900
           |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     841 EKMEAQGNAELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTE 900

Qy     901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     901 PLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATET 960

Qy     961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     961 SEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQA 1020

Qy    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1021 ILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAG 1080

Qy    1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1081 AEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETN 1140

Qy    1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1141 GSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQE 1200

Qy    1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1201 EHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQ 1260

Qy    1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1261 KAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA 1320

Qy    1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1321 APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAE 1380

Qy    1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1381 AVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPI 1440

Qy    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1441 IVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELE 1500

Qy    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKH 1560
           |||||||||||||||||||||||||||||||||||||| : ||||||||||||| |||||
Db    1501 SKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPAMQADSREPNRCWTKMKVAKMKH 1560

Qy    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
           ||||||||||||||||||||||||||||||||||||
Db    1561 PVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK 1596
```

```
RESULT 3
Q9WTQ5
ID   Q9WTQ5      PRELIMINARY;      PRT;  1684 AA.
AC   Q9WTQ5;
DT   01-NOV-1999 (TrEMBLrel. 12, Created)
DT   01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SSECKS (PKC binding protein SSECKS).
GN   AKAP12 OR SSECKS.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/C; TISSUE=TESTIS;
RA   Okita K., Kitamura H., Iwanaga T., Morimatsu M., Saito M.;
RT   "Isolation and characterization of mouse SSeCKS.";
RL   Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Camus A., Mesbah K., Rallu M., Babinet C., Barra J.;
RT   "Gene trap insertion into murine SSeCKS homologue interrupts one of
RT   the transcription units while revealing the existence of a second
RT   messenger exclusively express in the testis.";
RL   Submitted (DEC-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB020886; BAA76894.1; -.
DR   EMBL; AF326228; AAK16150.1; -.
DR   MGD; MGI:1932576; Akap12.
DR   InterPro; IPR001573; PkinA_anch.
SQ   SEQUENCE   1684 AA;  180694 MW;  E569D55762FCB19E CRC64;

  Query Match             80.7%;  Score 6512.5;  DB 11;  Length 1684;
  Best Local Similarity   83.9%;  Pred. No. 3.8e-252;
  Matches 1326;  Conservative   72;  Mismatches  154;  Indels   29;  Gaps   11;

Qy       1 MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSS 60
           |||||||||||||||| ||||||| ||||||||||||||||||||||||||||||||||:
Db       1 MGAGSSTEQRSPEQPAESDTPSELELSGHGPAAEASGAAGDPADADPATKLPQKNGQLSA 60

Qy      61 VNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDG 120
           ||||||| |||||||:|:||||||  |||||||||||:||||  ||||:|| ||||||| 
Db      61 VNGVAEQEDVHVQEESQDGQEEEVTVEDVGQRESEDVKEKDRAKEMAASSTVVEDITKDE 120

Qy     121 QEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180
           |||| ||||||||||:||||| | ||||||||||||||||||||||||||||||||||||
Db     121 QEETPEIIEQIPASESNVEEMAQAAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQL 180

Qy     181 LTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIP 240
           ||||||||||||||||||||||| ||| |||||||||||||||||||||||| |||||||
Db     181 LTVKKDEGEGAEASVGAGDHQEPGVET-VGESASKESELKQSTEKQEGTLKQAQSSTEIP 239

Qy     241 LQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTS 300
           ||||| |  ||||   :|||  :||||||   |||:|||::|||||||||||||||||||
Db     240 LQAESGQGTEEEAAKDGEENREKEPTKPLESPTSPVSNETTSSFKKFFTHGWAGWRKKTS 299
```

```
Qy      301 FKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPA-SEEQEPAEDTDQARLSADYEKVELP 359
Db      300 FKKPKEDDLETSEKRKEQEAEKVDEEEGEKTEPAPAEEQEPAEGTDQARLSADYEKVELP 359

Qy      360 LEDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGV 419
Db      360 LEDQVGDLEALS-EKCAPLATEVFDEKTEAHQEVVAEVHVSTVEKMTKGQ-GGAEVEGDV 417

Qy      420 VVEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHP 479
Db      418 VVEGSGESLPPEKLAETQEVPQEAEPVEELMKTKEVCVSGGDHTQLTDLSPEEKMLPKHP 477

Qy      480 EGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKR-GGGGDEEPGEYQHIH 538
Db      478 EGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGGDEEPGEYQHIQ 537

Qy      539 TESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWA 598
Db      538 TESPESADEQKGESSASSPEEPEEIACLEKGPSEAPQEAEAEEGATSDGEKKREGITPWA 597

Qy      599 SFKKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKR 658
Db      598 SFKKMVTPKKRVRRPSESDKEEELDKVKSATLSSTESTASGMQDEVRAVGEEQRSEEPKR 657

Qy      659 RVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEEASKDKEAGTDAVPAS 718
Db      658 RVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDGHRAEEEASKDKEA--DALPAS 715

Qy      719 TQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTE 778
Db      716 TQEQDQAHGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEERAEDSGAEQLASE 775

Qy      779 IEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAV 838
Db      776 IEPSREESWVSIKKFIPGRRKKRADGKQEQAAVEDSGPGEINEDDPDVPAVVPLSEYDAV 835

Qy      839 EREKME---AQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWIS 895
Db      836 EREKLEAQRAQENVELPQLKGAVYVSEELSKTLVHTVSVAVIDGTRAVTSAEERSPSWIS 895

Qy      896 ASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAV 955
Db      896 ASMTEPLEHAEGVATPPVGEVTEKDITAEATPALAQTLPGGKDAHDDIVTSEVDFTSEAV 955

Qy      956 TATETSEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEE 1015
Db      956 TAAETTEALRAEELTEASGAEETTDMVSAVSQLSDSPDTTEEATPVQEVEGGMLDTEEQE 1015

Qy     1016 RQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGP 1075
Db     1016 RQTQAVLQAVADKVKEDSQVPATQTLQRAGPKALEKVEEVEEDSEVLATEKEKDVVPEGP 1075

Qy     1076 VQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLT 1135
Db     1076 VQEAETEHLAQGSETVQATPESLEVPEVTEDVDRATTCQVIKHQQLMEQAVAPESSETLT 1135

Qy     1136 DSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNN 1195
Db     1136 DSETNGSTPLADSDTPNGTQQDETVDSQDSNAIAAVKQSQVTEEEAAAAQTEGPST-PSS 1194

Qy     1196 VPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKE-----EQE 1250
Db     1195 FPAQEEHREKPGRDVLEPT-QALAAGAVPILAKAEVGQEGEAGQFDGEKVKDGQCVKELE 1253

Qy     1251 VFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQ 1310
Db     1254 VPVHTGPNSQKTADLTRDSEVMEVARCQETESNEEQSISPEKREMGTDVEKEETETKTEQ 1313

Qy     1311 VSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQS 1370
Db     1314 ASEEHEQETAAPEHEGTHPKPVLTADMPHSERGKALGSLEGSPSLPDQDKADCIEVQVQS 1373

Qy     1371 LDTTVTQTAEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG 1430
Db     1374 SDTPVTQTTEAVKKVEETVATSEMDESLECAGAQSLPAEKLSETGGYGTLQHGEDTVPQG 1433

Qy     1431 PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVL 1490
Db     1434 PESQAESIPIIVTPAPESILHSDLQREVSASQKQRSDEDNKPDAGPDAAGKESAAREKIL 1493

Qy     1491 KAEPEILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCW 1550
Db     1494 RAEPEILELESKSNKIVQSVIQTAVDQFARTETAPETHASDLQNQVPVMQADSQGAQQML 1553

Qy     1551 TKMKDAKMKHPVPQPREDLQV 1571
Db     1554 DK-----------DESDLQV 1562
```

RESULT 4
Q99MP1

```
ID   Q99MP1      PRELIMINARY;      PRT;   1579 AA.
AC   Q99MP1;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Testis-specific SSeCKS.
GN   AKAP12.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=TESTIS;
RX   MEDLINE=21322683; PubMed=11429284;
RA   Camus A., Mesbah K., Rallu M., Babinet C., Barra J.;
RT   "Gene trap insertion reveals two open reading frames in the mouse
RT   SSeCKS gene: the form predominantly detected in the nervous system is
RT   suppressed by the insertion while the other, specific of the testis,
RT   remains expressed.";
RL   Mech. Dev. 105:79-91(2001).
DR   EMBL; AF326230; AAK16152.1; -.
DR   MGD; MGI:1932576; Akap12.
DR   InterPro; IPR001573; PkinA_anch.
SQ   SEQUENCE   1579 AA;  169743 MW;  D448047A8E6CE8F2 CRC64;

  Query Match             74.8%;  Score 6035.5;  DB 11;  Length 1579;
  Best Local Similarity   83.5%;  Pred. No. 3.9e-233;
  Matches 1232;  Conservative   67;  Mismatches  148;  Indels   29;  Gaps   11;

Qy      106 MAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKFVGFKF 165
Db        1 MAASSTVVEDITKDEQEETPEIIEQIPASESNVEEMAQAAESQANDVGFKKVFKFVGFKF 60

Qy      166 TVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEK 225
Db       61 TVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPGVET-VGESASKESELKQSTEK 119

Qy      226 QEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFK 285
Db      120 QEGTLKQAQSSTEIPLQAESGQGTEEEAAKDGEENREKEPTKPLESPTSPVSNETTSSFK 179

Qy      286 KFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPA-SEEQEPAEDT 344
Db      180 KFFTHGWAGWRKKTSFKKPKEDDLETSEKRKEQEAEKVDEEEGEKTEPAPAEEQEPAEGT 239

Qy      345 DQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEK 404
Db      240 DQARLSADYEKVELPLEDQVGDLEALS-EKCAPLATEVFDEKTEAHQEVVAEVHVSTVEK 298

Qy      405 TEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHTQ 464
Db      299 MTKGQ-GGAEVEGDVVVEGSGESLPPEKLAETQEVPQEAEPVEELMKTKEVCVSGGDHTQ 357
```

```
Qy    465 LTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKR- 523
Db    358 LTDLSPEEKMLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRG 417

Qy    524 GGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGT 583
Db    418 GGGGDEEPGEYQHIQTESPESADEQKGESSASSPEEPEEIACLEKGPSEAPQEAEAEEGA 477

Qy    584 TSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDE 643
Db    478 TSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELDKVKSATLSSTESTASGMQDE 537

Qy    644 VKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEE 703
Db    538 VRAVGEEQRSEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDGHRAEE 597

Qy    704 ASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKL 763
Db    598 ASKDKEA--DALPASTQEQDQAHGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKL 655

Qy    764 EEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDD 823
Db    656 EERAEDSGAEQLASEIEPSREESWVSIKKFIPGRRKKRADGKQEQAAVEDSGPGEINEDD 715

Qy    824 PNVPAVVPLSEYNAVEREKME---AQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGT 880
Db    716 PDVPAVVPLSEYDAVEREKLEAQRAQENVELPQLKGAVYVSEELSKTLVHTVSVAVIDGT 775

Qy    881 RAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAH 940
Db    776 RAVTSAEERSPSWISASMTEPLEHAEGVATPPVGEVTEKDITAEATPALAQTLPGGKDAH 835

Qy    941 DDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATP 1000
Db    836 DDIVTSEVDFTSEAVTAAETTEALRAEELTEASGAEETTDMVSAVSQLSDSPDTTEEATP 895

Qy   1001 VQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEEDSE 1060
Db    896 VQEVEGGMLDTEEQERQTQAVLQAVADKVKEDSQVPATQTLQRAGPKALEKVEEVEEDSE 955

Qy   1061 VLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQ 1120
Db    956 VLATEKEKDVVPEGPVQEAETEHLAQGSETVQATPESLEVPEVTEDVDRATTCQVIKHQQ 1015

Qy   1121 LMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQSQVTEEE 1180
Db   1016 LMEQAVAPESSETLTDSETNGSTPLADSDTPNGTQQDETVDSQDSNAIAAVKQSQVTEEE 1075

Qy   1181 AATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWL 1240
Db   1076 AAAAQTEGPST-PSSFPAQEEHREKPGRDVLEPT-QALAAGAVPILAKAEVGQEGEAGQF 1133

Qy   1241 DGEKVKE-----EQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEM 1295
Db   1134 DGEKVKDGQCVKELEVPVHTGPNSQKTADLTRDSEVMEVARCQETESNEEQSISPEKREM 1193

Qy   1296 ETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSL 1355
Db   1194 GTDVEKEETETKTEQASEEHEQETAAPEHEGTHPKPVLTADMPHSERGKALGSLEGSPSL 1253

Qy   1356 PDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGESPECVGAHLLPAEKSSATG 1415
Db   1254 PDQDKADCIEVQVQSSDTPVTQTTEAVKKVEETVATSEMDESLECAGAQSLPAEKLSETG 1313

Qy   1416 GHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAG 1475
Db   1314 GYGTLQHGEDTVPQGPESQAESIPIIVTPAPESILHSDLQREVSASQKQRSDEDNKPDAG 1373

Qy   1476 PDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQ 1535
Db   1374 PDAAGKESAAREKILRAEPEILELESKSNKIVQSVIQTAVDQFARTETAPETHASDLQNQ 1433

Qy   1536 VPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQV 1571

Db   1434 VPVMQADSQGAQQMLDK------------DESDLQV 1457
```

RESULT 5

```
Q9DDN8
ID   Q9DDN8      PRELIMINARY;      PRT;  2471 AA.
AC   Q9DDN8;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Gravin-like.
GN   GL.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21095185; PubMed=11165490;
RA   Klingbeil P., Frazzetto G., Bouwmeester T.;
RT   "Xgravin-like (Xgl), a novel putative a-kinase anchoring protein
RT   (AKAP) expressed during embryonic development in Xenopus.";
RL   Mech. Dev. 100:323-326(2001).
DR   EMBL; AF308810; AAG45105.1; -.
DR   InterPro; IPR001573; PkinA_anch.
SQ   SEQUENCE   2471 AA;  267987 MW;  1DCEFD7D443D1129 CRC64;

  Query Match             22.4%;  Score 1805;  DB 13;  Length 2471;
  Best Local Similarity   31.5%;  Pred. No. 4.5e-64;
  Matches  551;  Conservative  287;  Mismatches  539;  Indels  370;  Gaps   60;

Qy     82 EEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEM 141
Db     11 EEVSTPNEGPVEN--------MEVIPSEATPKTEIKEQGQDEAPEATN--ATSESPEEST 60

Qy    142 VQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQ 201
Db     61 EQANETQSNEVGFKKVFKFVGFKFTVKKDKVEKSEPVQLLTVKQEKVE----VNGTENHE 116

Qy    202 EPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQ 261
Db    117 KQN-------DVSGENVLE--TEDTKEVEKQSETDPIIQTSKTREEVSGEKPKSEEETKL 167

Qy    262 EKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEK------R 315
Db    168 EEEQVKTPESPTNPLVAETSSPLKKFFTQGWAGLRKKTSFRKSKEEDHQEVEKYIISEEK 227

Qy    316 KEQEAEKVDEEEKE-------------------------KTEPASEEQEPAEDTDQARL 349
Db    228 EKSDSQQVKEEDGENQIIVNGETLPKEANEIEESKACIDKETQPSAKELAKPEECPTDQL 287

Qy    350 SADYEKV--------ELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVH--- 398
Db    288 SVLKEDIKLENADICEVPVTESI---QTSAEAVHVSVPTEVSEQKKEVDTTEDNDVNKVT 344

Qy    399 -VSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKL---AEPQEVPQEAEPAEELMKSR- 453
Db    345 VVSTSTATSEE-----------VKEPLSEMAPDTEVICDKDNKDTDTKLAQTATLLESTE 393

Qy    454 -EMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSGLK 512
Db    394 VEICL----------------------EPIMTETELLSSQEKAKLQGSPLRKLFSGSGLR 431

Qy    513 KLSGKKQKGKRGGGGDEEPGEYQHIHTESPESAD--EQKGE-SSASSPEEPEETTCLEKG 569
Db    432 KLSGKKNKGKK---EDDTKAENT---TEVPVSSDIPESDGDNTSPSSPEDSAETSPTEKG 485

Qy    570 PLEAPQDGEAE-EGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSA 628
Db    486 AEDATQAVETEGDGATSDGERKREGITPWASFKKLVTPKKRPKRPSESDKEEEVEKIKTS 545
```

```
Qy    629 TLSSTDSTVS-EMQDEVKTVGEEQK----PEEPKRRVDTSVSWEALICVGSSKKRARKAS 683
          |:|||||  | | |:| |  ||||       || |::|| |||||||||||:|||||| |
Db    546 TMSSTDSAGSVENQEEAKENGEEQKLEKSTEENKKKVDNSVSWEALICVGSAKKRARKTS 605

Qy    684 SSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEGEGV 743
          ||||   : :  :     ||  |  | ||  ::    |:||::|     ||: : ||:||:|
Db    606 DSDDE-TQKNIDENKKIEEEVDKSKEHESEDPIVSSQEKEQVNDIPSPDQSTSPTEGDGG 664

Qy    744 STWESFKRLVTPRKKSKSKLEEKAED----SSVEQLSTEIEPSREESWVSIKKFIPGRRK 799
          |||:||||||| ||:||::: ||| |:     |: || ::: |   :|| ||| || ||||:|
Db    665 STWQSFKRLVIPRRKSRTRAEEKTEETTAVSNTEQPTSDGETGKEEGWVSFKKLIPGRKK 724

Qy    800 KRADGKQEQATVEDSGP----VEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLL 855
          |::||||| ||: |:|        |   ||| :||||||||||::| |:|| |||   :    :
Db    725 KKSDGKQEPATISDTGKSTEGCEAMEDDADVPAVVPLSEFDAAEQEKREAQQKEDTLNDV 784

Qy    856 G--AVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPV 913
             |   |  |: |:| ::| |::| |||||:|:|:|||||||:||| :|       |
Db    785 SKDEVKTPEVNSEGLIHAITVTVVEGERAVTSLEDRAPSWISANVTETVEQANELTEPNT 844

Qy    914 EEVTEKDIIAEETPV---LTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVT 970
          :|  : :|   ||  :    ::| :  |      : : :||: ||||:||           ||
Db    845 KERIKSEITVEEAFIFGEVSQMVGSG-----NTIINEVELTSEALTA--------LEEAI 891

Qy    971 EASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK 1030
          | |  |||||:|:|| ||| ||      || ||: | :  |      :   |:|    :|    | ::
Db    892 EYSCAEETTEMISAASQLDDSFTPAEEVTPIPEEDEGSQSLDMPEKQMDPVLDE-ATELT 950

Qy   1031 EES-----QVPATQTVQRTGSKALEKVEEVE-----------EDSEVLASEKEKDVMP-- 1072
          |:|                                        |:|  :  : |: :: |
Db    951 EQSILSLDKIQLNESISTSAEEHDTKVDKIEEANICISSKQDEESTIFSDERIEECTPVL 1010

Qy   1073 ---------KGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHV---ATCQVIKLQQ 1120
                   | |: ||                 :|: |     |:|:    |   :::| :
Db   1011 AERHDPSELKSVVESNGA--------------DSVSV------VEHIIEDAVIELVKERY 1050

Qy   1121 L-MEQA---VAPESSETLTDS--------ETN-GSTPLADSDT---ADGTQQDETIDSQD 1164
          |  :||        :: |  ||:          ||: ||  : ||:|   |  :| |: :|:
Db   1051 LSSDQADIYAQVDTKEHPTDAVLLEVNREETSVGSVEVQDSETILILDESQADKIVDTAQ 1110

Qy   1165 SKATAAVRQSQVTEEEAATAQKEEP---------------STLPNNVPAQEEHGEE--PG 1207
            :: : :  |:    || |      ::                    ||  : :| :|:  :
Db   1111 KQSESILVLSEEKVEECAIMSEDHKTASVHPKCQFEVTSVSTEDDALPVKEQVSDNSISA 1170

Qy   1208 RDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQ------------------ 1249
           :::    ::       ::| | | :|  |  :   |  :   :|||
Db   1171 VKLVDVAAEQTVEKSIPTLEQLKELSEEQSSTISVSTVHEEQVKEAALISDEQDDALPVK 1230

Qy   1250 EVFVHSGPNSQKAADVTYDSEV--------------------MGVAGCQEKESTEVQSLS 1289
          |  : :  :: |  || :  |                     : |:    |::  |    :|
Db   1231 EQVLDNSISAVKLVDVAAEQTVEKSIPTLEQLKELSEEQSSSISVSTVHEEQVKEAALIS 1290

Qy   1290 LEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSL 1349
           |:  :|      | ||| |  :||  :| :  |    |  |    |      :  :|  :
Db   1291 DEQETIEHPHASEDVETK-EGAAEEQVEEKSLIYEEVTESAQVEECSLVFAETQRK---- 1345

Qy   1350 GGSPSLPDQD---KAGCIEVQVQSLDTTVTQTAEAVEKVI-----ETVVISETGESPECV 1401
            | |: :::   |:  :|   ||   |:   :  |: |       ||    : | | :
Db   1346 -ESVSILEEELARKSDPVEPLEQS-GITLATVEDTTERAIFADKEETAKSVKIGTSAKAA 1403

Qy   1402 GAHLLPAEK------SSATGGHWT----LQHAEDTVPLGPESQAESIPIIVTPAPESTLH 1451
               |   :|          |      :       |  |           :: |    :   :| :| |:|
Db   1404 LNALFENDKHSLEDVSSEVGKQVSECSAVMFDEQAVEAAQSTEKEQ--VAKSPTPVATA- 1460

Qy   1452 PDLQGEISASQRERSEEE------------------------------DKPDAGPDADG 1480
            |  ||::   | :||                               :| |    |
Db   1461 --LTEEIASLVEEAVKEEKNAEYPHSETQISEIPSSLPIDQTKEVTTLAEKYDTEIILDE 1518

Qy   1481 KESTAIE--KVLKAEPE------ILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDS 1532
            |  |:|   ||  |  |        ::| |    :   :  | |||              | |  :
Db   1519 TEKQAMESVSVLSGEQEVENFNSVVEEEVSKSAFIAGVEQTAESW---------TTASEE 1569

Qy   1533 QTQVPACRLDSREPNRCWTKMKDAKMKH----PVPQPREDLQVLTVLEAWAQPRKCLPRL 1588
             ||:               | |    | :          | :| |:     | ||  |:        :
Db   1570 QTR------------ECTTTADDVQAVECNAITVDEPSEEKPSATSLEHSAE----TVSV 1613

Qy   1589 QLKAPVS 1595
          |: |::
Db   1614 SLECPIT 1620
```

RESULT 6
Q9VC00

```
ID   Q9VC00      PRELIMINARY;      PRT;  2768 AA.
AC   Q9VC00;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   CG13648 protein.
GN   CG13648.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003750; AAF56376.1; -.
DR   FlyBase; FBgn0039257; CG13648.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00093; vwc; 1.
DR   SMART; SM00214; VWC; 4.
DR   PROSITE; PS01208; VWFC; 1.
SQ   SEQUENCE   2768 AA;  294032 MW;  CA929A21774E4684 CRC64;

  Query Match             8.0%;  Score 642;  DB 5;  Length 2768;
```

```
Best Local Similarity   22.2%;  Pred. No. 1.3e-17;
Matches  444;  Conservative  250;  Mismatches  735;  Indels  572;  Gaps   94;

Qy       2 GAGSSTEQRSPEQPAGSDTPSELVLSGHGPA------AEASGAAGDPA-DADPATKLPQK 54
Db     550 GATSASEE-SEEQDEGKSTEAPTSVDDIEPAKPTESSEEASGEGEDVAKETTPAGE---- 604

Qy      55 NGQLSSVNGVAE--QGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTA 112
Db     605 ----ASIAGEEEIVKGTTPAGEPSSEG-DEEIVKGTTPAEESSSESE----DELTKVTTP 655

Qy     113 VEDITKDGQEETSEIIEQIPASENNV---EEMVQ---PAESQANDVGFKKVFKFVGFKFT 166
Db     656 AGEPSVAGEEEIAK--ETTPAGEPSIAGEEEIVKVTTPA-GESSIAGEEEI-----VKVT 707

Qy     167 VKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGE--SASKESELKQSTE 224
Db     708 TPAGESSSEGEEEIIKVTTPAGE----SSSEGDEEIVKESTPAGEPISEGEEDVIKATTS 763

Qy     225 KQEGTL---KQEQSSTEIPLQAESDQAAE----EEAKDE------------GEEK---- 260
Db     764 APKSDIEGVKEPETATEVPAEEVEDFAKPTTPIAEAEEEPIAGTPIPTDGISGEEEIVKG 823

Qy     261 -----QEKEPTKSPESPSSPV-----NSETTSSFKKFFTHGWAGWRKKTSFK-------- 302
Db     824 TTPQTLEEQPEISEESTEVPVAEDDLSSSTSASAIASSTEGVQDAASETTSSAPARAGDK 883

Qy     303 ----------KSKEDD--------------------LETAEKRKEQ---EAE------- 321
Db     884 DEAATTVPTAQDKDDEEVEQDATDLPVEDVVQSTTAKTTTTEQPKEESSTEAEDAEIEVT 943

Qy     322 ---KVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVG----DLEA----- 369
Db     944 TSSPADKQEVPEAEPADKDHKDEEDVQTA--------TDLPIKSDIGPPVVDTEATTGQP 995

Qy     370 -SSEEKCA---------PLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQG--------- 410
Db     996 ETSDETATDKPPSVYLPPVSQEVPSSTAKVDNRNDFETEKPTLPPSGEDQSSEPLPAMDL 1055

Qy     411 -GGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAE-----------------ELMKS 452
Db    1056 PAGIPGEGDCLVEG-------KTYANNTIVPATA-PCDVSCKCISSLVACQQMECKLPEN 1107

Qy     453 REMCVSGGD------HTQLTDLSPE-----EKTLPKHPEGIVSEVEMLSSQERIK----- 496
Db    1108 LEKCTVAADLLDGCCPTYICDESTESAEKDEESTAKPDNKIDEDVSEISTEEIPKDVIMP 1167

Qy     497 --VQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSA 554
Db    1168 TGITEQPLSHV-----------------------KPDEEIQPVTSVPAQFDE---STTA 1200

Qy     555 SSPEEPEETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPS 614
Db    1201 KVDKKPIDESAEDKKPI-----GESEEDSKPIDESEED--------KKPVEESAEDKKPV 1247

Qy     615 ESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKP-EEPKRRVDTSVSWEALICVG 673
Db    1248 E-DSEEKEKPLPTVIPASEIEKESKPEDEKKTEADFAAPTEQPEATTPAQIADTAEKEVD 1306

Qy     674 SSKKRARKASSSDDEGGPRTLGGDSHRAEEAS-KDKE--AGTDAVPASTQ---------- 720
Db    1307 DKLATTSAPVSGEDELKP---ADEKKRTETAQIPDAEIPASTDEPESSTELPTVDLDKKP 1363

Qy     721 EQDQAQGSSSPE-------PAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVE 773
Db    1364 EEDSTKGTEAPESDKVPEVPTSASTENE-IEESDKFTTVAPPKISASDETEPTAEEDLV- 1421

Qy     774 QLSTEIEPSREESWVSIKK----------FIPGRRKKR---------AD-GKQEQATVED 813
Db    1422 --PATFEPIESEFEVSTKKPAVQGPPLPTLAPAQPEKKPVDAETSTEADISTEPSAEVEK 1479

Qy     814 SGPVEINEDDPNV-------PAVVPLSEYNAVEREK-MEA-------------QGNTELP 852
Db    1480 EASGETSESDNEIDAGASSTPVPVSADEDKTPSTEKTVEADDKFTTVAPLAGDEEESNLP 1539

Qy     853 QLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEER----------------------- 889
Db    1540 KLPQDIF--EEEAPVAVTTAAPSKDDGEQKPVEVEEKPIEDGQKPIEDETSTPTSSENEI 1597

Qy     890 -----SPSWISASVTEPLEHTAG-----EAMPPVEEVTEKDIIAEETP---VLTQTLPEG 936
Db    1598 EPESDRATTIAPSKEEPSEPSTGAPTKDEPAEPSTDAPESD-ESKETPESEVPTTVAPAG 1656

Qy     937 KDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTE------ASGAEE---TTDMVSA--- 984
Db    1657 EKIPTSSITPDEEPTATSAPVAKPDEDVEKETSTEIPTDAPASSEEDENSSTDQIPSEVP 1716

Qy     985 ---------------VSQLTDSPDTTEEATPVQEV-------------ESGVLDTEE--- 1013
Db    1717 EKKPETPAQTPEEGDIVGATAAPTTSDEVPPVQRLPEEVLAEIPQPSTETGIKQQDETTA 1776

Qy    1014 ---------------EERQTQAILQAVADKVKEESQVPATQTVQRTGSKALEKVEEVEED 1058
Db    1777 APSIDRKEPYVTEIDEEATTVAPISEKDEKPTEE-------------EKPVEQKPTGEEP 1823

Qy    1059 SEVLASEK--EKDVMPKGPVQ-EAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQV 1115
Db    1824 SEEEEKEKPIEQDVSTEGPVSTEASEAGSTESSEEVKPSTEG-EVAEKPEDKQPSSTAQA 1882

Qy    1116 IKLQQLMEQAVAPESSETL--TDSETNGSTPLADSDTADGTQQDETIDS------QDSKA 1167
Db    1883 -------PVETIPEISTELPAQDGDKPTSEAPVDSDEDTSAPSDEKIPSVSGEEVEGPEV 1935

Qy    1168 TAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLA 1227
Db    1936 TTASPQA-AEEDELKTPAESEPSS-TDKVPETEYQKPE------DETKADETPESVTQVS 1987

Qy    1228 KTEVGQEGEVDWLDGEK--------VKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQE 1279
Db    1988 DVATSTSAPVAGGDIEKDEQATTASPEEEEEI----KPTIAPAAEIPQPS---------E 2034

Qy    1280 KESTEVQSLSLEEGEMETDVEKEKR---ETKPEQVSEEGEQETAAPEHEG---------- 1326
Db    2035 KEPVDEQ--EVESGTKATPAESDGQPIDEIAPATSGPIDEASTAAPTKEESTTVASAASP 2092

Qy    1327 ----------TYGKPVL---TLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDT 1373
Db    2093 AVHDDEIKDVTTTQPVADEKEVAAPQDETKTSIDVSTDSPTAQDDEKQDKTEAPV--APT 2150

Qy    1374 TV-----------TQTAEAVEKV-IETV----VISETGESPECVGAHLLPAEKSSATGG 1416
Db    2151 TVSSPTADSAADSSTPTVEVPSPVEIDTKPMDDIMSQTIAPHTADGAASTSTEDEDQAPV 2210

Qy    1417 HWTLQHAEDTVPLGPESQ------------AESIPIIVTPAPE----STLHPDLQGEIS 1459
Db    2211 TVSPQDAEKT-PVSPAPQDSDKTPSSEAPQDADEIPATATPLDDNKIPATVAPQTDDGVP 2269

Qy    1460 ASQRERSEEEDKPDAGP-DADGKESTAI----EKV-LKAEPEILELESKSNKIVLNVIQT 1513
Db    2270 ATAAPLDEDKIQTTAAPLDEEKIPSTAAPLDDEKIPAPVSPVVFDVEPSSEK-------P 2322

Qy    1514 AVDQFARTE-TAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVL 1572
Db    2323 AVSEYDGEESTEPPVHDVETSTDEPT---------------SDAKLKPPTSAPATPSE-S 2366

Qy    1573 TVLEAWAQPRKCLPRLQLKAP 1593
Db    2367 PATEAEIVPETAAPELEKEVP 2387
```

RESULT 7
Q9U459

```
ID   Q9U459      PRELIMINARY;      PRT;   5458 AA.
AC   Q9U459; Q9GPG4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
```

```
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Erythrocyte membrane-associated giant protein antigen 332.
OS   Plasmodium falciparum.
OC   Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX   NCBI_TaxID=5833;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=FCC1/HN;
RA   Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;
RT   "Molecular cloning and structure analysis of the Plasmodium falciparum
RT   erythrocyte membrane-associated giant protein Ag332 (Pf332) gene.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF202180; AAF15293.3; -.
DR   InterPro; IPR001313; Pumilio/Puf.
SQ   SEQUENCE   5458 AA;  615269 MW;  4F6A37D92BD80172 CRC64;

  Query Match             7.9%;  Score 635;  DB 5;  Length 5458;
  Best Local Similarity   23.4%;  Pred. No. 5.6e-17;
  Matches  411;  Conservative  292;  Mismatches  688;  Indels  366;  Gaps   79;

Qy        4 GSSTEQRSPEQPA--GSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQLSSV 61
Db      977 GSVTEQLVEEEKGLINEDNEKEELIT--EMSEEIKSVNEEIEETDLSTEEIIKQQGLATY 1034

Qy       62 NGVAEQ---GDVHVQEEN---QEGQEE----EVVDEDVGQRESEDVREKDRVEEMAANST 111
Db     1035 EFVEEEKSLTDKLLEEESVTKEVGETELSTQEVVDEKVSVTEEVIEEEKSVSEEVLEEGS 1094

Qy      112 AVEDITKDGQ---------EETSEIIEQIPASE------------------------- 135
Db     1095 ATEEVVEEGSCTEIVVEKGSDTEIVVEEGSATEIVVEEGSATEVVVEEGSVSEEMLEEEG 1154

Qy      136 NNVEEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKD--------E 187
Db     1155 SATEEVVEEGLSSDNVQKSKGVIENVGEIYSVKTAKDESMN--EKIPLEKSSFVDDESFK 1212

Qy      188 GEGAEASVGAGD--------HQEPSVETAV---------------GESASKESEL--KQ 221
Db     1213 GQGPTDNVSVEDVNSEDIINEHTPLEETKIEELPTEYITTADIHTRGETETKYNLIYEKI 1272

Qy      222 STEKQEGTLKQEQSSTEIPLQAES---DQAAEEEAKDEGEEKQ----EKEPTKSPESPSS 274
Db     1273 NEEVEKAKFQEEKITENIPVERESVTEDIVQEPSLAQEVEQKESDTNEIEETKLANEKII 1332

Qy      275 PVNSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEA----EKVDEEEKEK 330
Db     1333 PEVSVTENVVEK---------EGLDTEEVLEEDESITEEIVEEEVSACEEIVEEEESSS 1382

Qy      331 TEPASEEQEPAEDTDQARLSADYEKV---ELPLEDQVGDLEASSEEKCAPLAT--EVFDE 385
Db     1383 EEIVEEEESSSEEIVEEEESSSEEIVEEEESSSEEIVEEEESSSEEIIEEVSSTEEVLEE 1442

Qy      386 KMEAHQEVVAEVHVSTVEKTE---------EEQGGGGE----AEGGVVVEGTGESLPPE 431
Db     1443 EGSVTEEIVEE-EVSTTEEVKDIGSVSEEVLEEEGFGTEEFVGQQGSVIEEIVETESSTE 1501

Qy      432 KLAEP-----QEVPQEAEP-AEELMKSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSE 485
Db     1502 KVLEDVGSNVEEIVQEEGPVAQEIVHEE---VSTTEKHDEVDRSTTEEIVEK--VGSVSE 1556

Qy      486 ---VEMLSSQERIKVQGSPLKKLFSSSGL---------------KKLSGKKQKGKRGGG 526
Db     1557 EIIVEEVSASEEIVEEGSVTEEVVEEEKLINEVGETESVTEEIVQKEVSDAEEVLGQEGS 1616

Qy      527 GDEEPGEYQHIHTE--SPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAEEGTT 584
Db     1617 MNEEILEKESIVEEIVGPEGSVTEEIVDYGSFAEEVKEEELVTE---EAVQ----YEGSV 1669

Qy      585 SDGEKKREGIT-----PWASFKKMVTPKKRVRRPSESD--KEEELEKVKSAT-LSSTDST 636
Db     1670 TEEIKEEESITENEAIEESAFAEIIEEK----GPNNDEIVKEEGLDTEEIVNEVSVTDEV 1725

Qy      637 VSE---MQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSS---KKRARKASSSD---D 687
Db     1726 IEEEKLVNEQI--VGEERSVTEKPVEVERSAT-EDLVEEGASVTEKVSVHEGSTTEQILD 1782

Qy      688 EGGPRTLGGDSHRAEEASKDKEAGTDAV---PASTQEQDQAQGSSSPEP------AGSPS 738
Db     1783 ESVAEEIVEEEVSVDDKIIEEEVSVDEVVEEEGSVIEEIVEEEESVPEEILEEELSGSEE 1842

Qy      739 -----------EGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESW 787
Db     1843 VLEDEWVTDAFMGQEGSVIEEIEEIV---DGEGSITEEIVEDGSANEKIVEEEPSRVEEV 1899

Qy      788 VSIKKFIPGRRKKRADGKQEQATVEDSGPV-EINEDDPNVPAVVPLSEYNAVEREKMEAQ 846
Db     1900 LGKEGFVIEEIIEEGSVIEQ---VEDTKTVSEKSEESSAIEEVKEVKEEESISEKIVEKE 1956

Qy      847 GNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTA 906
Db     1957 ES-----------VTEEIVRQEESTTEKIVKDVSPTEDFVEQ------TDSVTEKVIEQE 1999

Qy      907 GEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRT 966
Db     2000 GSNTEVAEDVEEKESASDE-----------HEQEDVSVNAQVTCEKKSVTKEIVDEVSRT 2048

Qy      967 EEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVA 1026
Db     2049 EEIVEENGS--VTEGVDETGSVTE--EIIEEATVTEEVVEDGSVTEEVVEDGSVIQEVVE 2104

Qy     1027 D-KVKEESQVPATQTVQRTGSKALEKV-------EEVEEDSEVLASEKEKDVMPKGPVQE 1078
Db     2105 DGSVTEE-------IVQENGSVTEEIVEEEGSVNEEVEEEVSVSGEVDETEYVTE-EVEE 2156

Qy     1079 AGA--EHLAQGSETGQATPESLE-----VPEVTADVDHVATCQVIKLQQLMEQAVAPESS 1131
Db     2157 EGSVVEEIVE--EEGSVVEEIVEEEGSVVEEIVEEEGSVVEEIVEEEGSVVEEIVEEEGS 2214

Qy     1132 ETLTDSETNGSTPLADSDTADGTQQDETIDSQDSKATAAVRQ-----SQVTEEEAATAQ- 1185
Db     2215 VVEEIVEEEGSV-VEEIVEEEGSVVEEIVEEEGSVVEEIVEEEGSVVEEIVEEEGSVVEE 2273

Qy     1186 --KEEPSTLPNNVPAQ----EEHGEEPGRDVLEPTQQELTAAAVPVLAKTE-------VG 1232
Db     2274 IVEEEGSVVEEIVEEEGSVVEEIVEEEGSVVEEIVEEE--GSVVEEIVEEEGSVVEEIVE 2331

Qy     1233 QEGEVDWLDGEKVKEEQEV----FVHSGPNSQKAAD-VTYDSEVMGVAGCQEKESTEVQS 1287
Db     2332 EEGSV-----SEVVDETELVNDEIVEQAPFTEEVEEQVSVNDEIIEDASVAE-------- 2378

Qy     1288 LSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPE---HEGTYGKPVLTLDMPSSERGK 1344
Db     2379 -AVEESESITESVSQEEETEKGFVIEKVEETGAVTEEIVQDGLITEEI--LEESESVNGE 2435

Qy     1345 ALGSLGGSPSLPDQD--------KAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGE 1396
Db     2436 IINKESDAEEILETEFLTEEVVGQAGSTSEEIVEEEGSVTKEVEEKESVTEELV-DEGSV 2494

Qy     1397 SPECVGAHLLPAEKSSAT------GGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTL 1450
Db     2495 TEELV-------DEGSVTEEVVEQGGSIAQEIVEE------ESATEEI-----IRDETNV 2536

Qy     1451 HPDLQGEISASQRERSEEEDKPDAGP-DADGKESTAIEKVLKAEPEILELESKSNKIVLN 1509
Db     2537 EEVLEKEGSA-----TEEIVQDGSGTNDFVGKQGSVIEEVV--EEEISTTEEK-----LK 2584

Qy     1510 VIQTAVDQFARTETAPE 1526
Db     2585 EEASAIEEFVEEESIRE 2601


RESULT 8
Q95YM2
ID   Q95YM2      PRELIMINARY;      PRT; 17352 AA.
AC   Q95YM2;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
```

```
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   I-connectin.
GN   I-CON.
OS   Procambarus clarkii (Red swamp crayfish).
OC   Eukaryota; Metazoa; Arthropoda; Crustacea; Malacostraca;
OC   Eumalacostraca; Eucarida; Decapoda; Pleocyemata; Astacidea;
OC   Astacoidea; Cambaridae; Procambarus.
OX   NCBI_TaxID=6728;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21423462; PubMed=11532946;
RA   Fukuzawa A., Shimamura J., Takemori S., Kanzawa N., Yamaguchi M.,
RA   Sun P., Maruyama K., Kimura S.;
RT   "Invertebrate connectin spans as much as 3.5 micrometer in the giant
RT   sarcomeres of crayfish claw muscle.";
RL   EMBO J. 20:4826-4835(2001).
DR   EMBL; AB055861; BAB64297.1; -.
DR   InterPro; IPR000577; FGGY_kin.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 48.
DR   Pfam; PF00018; SH3; 1.
DR   PROSITE; PS00933; FGGY_KINASES_1; UNKNOWN_3.
DR   PROSITE; PS50002; SH3; 1.
SQ   SEQUENCE   17352 AA;  1962348 MW;  4BA157BEC042E42D CRC64;

  Query Match             7.3%;  Score 588;  DB 5;  Length 17352;
  Best Local Similarity   22.3%;  Pred. No. 1.6e-14;
  Matches  381;  Conservative  251;  Mismatches  637;  Indels  438;  Gaps   76;

Qy       41 DPADADPATK------LPQKNGQLSSVN--------------------GVAEQGDVHVQE 74
Db    14336 DLADEKPKTEIAKSLLLPQDSIAVQEVSVKEAPVDIKELKPKSETAKPVVSTQESLSVSE 14395

Qy       75 ENQEGQEEEVV-DEDVGQRESEDVREKDRVEEMAANS-------------------TAVE 114
Db    14396 VQVEGQTQEMVTPKQISETARVSVTKKDSVQVLEVSSELAPDDIPETRLIEDKAKVTVTE 14455

Qy      115 DITKDGQEETSEIIEQIPASENNVEEMVQPAESQ-ANDVGFKKVFKFV--GFKFTVKKDK 171
Db    14456 NLQRADQEDV--IVLRGPTKEKKIIEVVEESDEELVKEIPEVKEEEIVEEQFKLKPKTKK 14513

Qy      172 NEKSDTVQLLTVKKDE--GEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGT 229
Db    14514 KVKEEITEEVIIKKPEEKEEVIEEVIKKPDEIEVTEEVMIKKKPKKKTEVTTEEVVEEVT 14573

Qy      230 LKQEQSSTEIPL---QAESDQAAEEEAKDEGEEKQ----------EKEPTKSPESPSSPV 276
Db    14574 IKKPEVIEEVTVKKPEEKIEEITEEVTIKKPEEKVVEEITEEVQIKKKPKKKPEPKIEEV 14633

Qy      277 NSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEE------EKEK 330
Db    14634 TEEVT--------------LKKPEEKIAEEVTEEIVVKKAEEKVEEVTEEVQIKRKPKKK 14679

Qy      331 TEPASEE-------QEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVF 383
Db    14680 PEPKIEEVTEEVTIKKPEEVTEEVVIKKPEEVVHEEVSEEV--TIKKPEEKVVEEVTEEV 14737

Qy      384 DEKMEAH-------QEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGE---SLPPEKL 433
Db    14738 QIKKKPKKTPEPKIEEVTEEV---TLKKPEEK----------VVEEVTEEVKIKRKPKKK 14784

Qy      434 AEP--QEVPQEA---EPAEELMK--SREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEV 486
Db    14785 PEPKIEEVTEEVTIKKPEEKVIEEVTEEVVIKKPEEKVHEEVS-EEVTIKKPEEKIVEEV 14843

Qy      487 EMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPE-SA 545
Db    14844 -----TEEVKIKRKPKKK--------------------------------------PEPKI 14861

Qy      546 DEQKGESSASSPEEP--EETTCLEKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKM 603
Db    14862 EEVTEEVTLKKPEEKVIEEVT--EEVVIRKPEEKVVEE---------------------- 14897

Qy      604 VTPKKRVRRPSESDKEEELE--KVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVD 661
Db    14898 VSEEVTIKKPEEKVVEEVTEEVKIKMKPKKKPEPKIEEVTEEV-TI---KKPEE---KID 14950

Qy      662 TSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQE 721
Db    14951 EEITEEVQIKRKPKKKPEPKIEEVTEEVTIK---------KPKEKVVEEITEEVVIKKPE 15001

Qy      722 QDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEP 781
Db    15002 EKVPEEVSEEVTIKKPEEKVVEEVTEEVKIKMKLKKKPEPKIEEVTEEVTIKKPEEKIDE 15061

Qy      782 SREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVER- 840
Db    15062 EITEE-VQIKR----KPKKKPEPKIEEVT------EEVTLRKPKEEAIEEITEEVTIKKP 15110

Qy      841 -EKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVT 899
Db    15111 EEKIE-----EVSEEVTIKKPEEKVVEVVTEEVKIKMKPKKKPEFKVEEVSE---EITIK 15162

Qy      900 EPLEHTAGE---------AMPPVEEVTE-------KDIIAEETPVLTQTLPEGKDAHDDM 943
Db    15163 KPEEEVAEEVQIRKPKKKSVFKTEEISEEITLQKPKEKVEEVTEEVTLKRPEEK------ 15216

Qy      944 VTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQE 1003
Db    15217 IVEEV---SEEVTIKKPDEKV-VEEVTE----EVQIKMKPKMKPKYEVEEVTEEMT-IKQ 15267

Qy     1004 VESGVLDTEEEERQTQAILQAVADKVKEE---SQVPATQTVQRTGSKALEKVEEVEEDSE 1060
Db    15268 PE------EEVEEITEEVRFKLKPKKKPEPKYEEVSSEVTIMR---KREDKVEEVVEDVH 15318

Qy     1061 VLASEKEKDVMPKGPVQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQ 1120
Db    15319 IKMKPKEK---PQEKFDDVSEEITISKPETKEQVTE-----EITEDVQ-------IKLK- 15362

Qy     1121 LMEQAVAPESSETLTDSETNGSTPLADSDTADGTQQD--ETID-SQDSKATAAVRQSQVT 1177
Db    15363 -------PKKKPDIRTEEVTEEVTIIRKEEPSEEEEDISETVQIKRKPKRKITTKEEEVT 15415

Qy     1178 EEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEV 1237
Db    15416 E---VTIQKPKPK-------------EEPVEEVTEEFMLKRPKPKKPEEPQQEVSEEVTI 15459

Qy     1238 DWLDGE-KVKEE--QEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGE 1294
Db    15460 KQESPEPTVTEEVTEEVSFRVKPKRKKPIVTEEISEEFTLRKPKEPE----------EEGP 15510

Qy     1295 METDVEKE-KRETKP-----EQVSEEGEQETAAPE---HEGTYGKPVLTLDMPSSERGKA 1345
Db    15511 YEVSEEVTIKRKPKPFIRETQEISEEVEMKLTEPEDFTHEETTEDVKIKMKRPKKK---- 15566

Qy     1346 LGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETG-ESPECVGAH 1404
Db    15567 --------------------------YSVTE-----ETIDSKMVIQESDEEEPEVIYEE 15594

Qy     1405 L--LPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQ 1462
Db    15595 LEDVVYQKKIIKDG---TEIAQEFVQLRKKKPKKSVTV--------------------- 15629

Qy     1463 RERSEE--------EDKPDAGP-DADGKESTAI-----EKVLKAEPEILELESKSNKIVL 1508
Db    15630 RDEGEETVVIKKSITDETDAGPKHEDIEESFQLRLPRKKKYKVEEEEEVEVGVKIRREPT 15689

Qy     1509 NVIQTAVDQFARTET-APETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPRE 1567
Db    15690 FEVREDVEESFRIKTMEPE-----------------EEPEEEYEE-RDVEGKFTFDLPRK 15731

Qy     1568 DLQVLTVLEAWAQPRKCLPRLQLKAPV 1594
```

```
                :|  |     : |  | || |  : ||
Db  15732 KIQKRTEEIEYTQKVK-LPPLPKQPPV 15757


RESULT 9
Q22248
ID   Q22248      PRELIMINARY;      PRT;  1829 AA.
AC   Q22248;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   T06D8.1 protein.
GN   T06D8.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Palmer S.;
RL   Submitted (APR-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99069613; PubMed=9851916;
RA   none;
RT   "Genome sequence of the nematode C.elegans: A platform for
RT   investigating biology.";
RL   Science 282:2012-2018(1998).
DR   EMBL; Z49130; CAA88964.1; -.
DR   InterPro; IPR003014; PAN.
DR   Pfam; PF00024; PAN; 3.
SQ   SEQUENCE   1829 AA;  187218 MW;  18A4564A71BBC3CE CRC64;

  Query Match             7.3%;  Score 585.5;  DB 5;  Length 1829;
  Best Local Similarity   23.2%;  Pred. No. 1.5e-15;
  Matches  375;  Conservative  174;  Mismatches  662;  Indels  405;  Gaps   65;

Qy      2 GAGSSTE-----QRSPEQPAGSDTPSELVLSGHGPAAEASG-----AAGDPADADPAT-- 49
Db    106 GSGEETTVVAVVESSGEEPASSSTSVPTELSKDDQVTEASGEETTTAAATEASSEETTTS 165

Qy     50 KLPQKNGQLSSVNGVAE-------------QGDVHVQEENQEGQEEE----VVDEDVGQ 91
Db    166 AVTEGSGEETTTSAVTEASSEATTTPAGTEASGEETTTSAVTEGSGEETTVVAVVESSGE 225

Qy     92 RESE-------DVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMV-- 142
Db    226 EPASSSTSIPTELSKNDQVTEASGEETITAAATEASEETTTSAVTEGSGEDTTVVAVVEL 285

Qy    143 ---QPAESQANDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAG- 198
Db    286 SGEQPASSS-----------------TSIPTELSKDDQV-------TEASGEETTAAAT 321

Qy    199 DHQEPSVETAVGESASKESELKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGE 258
Db    322 EASEETTTSAVTEGSGEETTVVAVVESSGE--EPASSSTSIPTELSKDDQVTEAS---GE 376

Qy    259 EKQEKEPTK-SPESPSSPV---NSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEK 314
Db    377 ETTTAAATEASEETTTSAVTEGSGEDTTVVAVVESSGEQPASSSTSIPTELSKDDQVTEA 436

Qy    315 RKEQEAEKVDEEEKEKTEPASEEQEPAEDT----------DQARLSADYEKVELPLEDQV 364
Db    437 SGEETTTAAATEASEETTTSAVTEGSGEDTTVVAVVESSGEQPASSSTSIPTELSKDDQV 496

Qy    365 GDLEASSEEKCAPLATEVFDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGT 424
Db    497 --TEASGEETTTAAATEASEE-------------TTTSAVTE---GSGEETTVVAVVESS 538

Qy    425 GESLPPEKLAEPQEVPQEAEPAEELMKSREMCVSGGDHT---QLTDLSPEEKTLPKHPEG 481
Db    539 GE----EPASSSTSIP------TELSKDDKVTEASGEETTTAAATDASSEETTTSAVTEG 588

Qy    482 IVSEVEMLSSQERIKVQGSPLKKLFSSSGLKKLSGKKQKGKRGGGDEEPGEYQ-HIHTE 540
Db    589 SGEETTVVAVVE-------------------------------SSDEEPASSSTSIPTE 616

Qy    541 SPESADEQKGESS------ASSPEEPEETTCLEKGPLEAPQDGEAEEGTT-----SDGEK 589
Db    617 L--SKDDQVTEASGEETTTAAATEASEETT------TSAVTEGSGEETTVVAVVESSGEE 668

Qy    590 KREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVKSATLSSTDSTVSEMQDEVKTVGE 649
Db    669 ------PASSSTSIPT------ELSKDDKVTEASGEETTTAAATDASSEETTTSAVTEGS 716

Qy    650 ----------EQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSH 699
Db    717 GEETTVVAVVESSDEEPASS-STSIPTEL---------------SKDDQVTEASGEETT 759

Qy    700 RAEEASKDKEAGTDAVPASTQEQDQAQG---SSSPEPAGSPSEGEGVSTWESFKRLVTPR 756
Db    760 TAAATEASEETTTSAVTEGSGEETTVVAVVESSGEEPASSSTS-------------IPT 805

Qy    757 KKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGP 816
Db    806 ELSKD--DQVTEASGEETTTAAATEASEETTTS------------------AVTEGSG- 843

Qy    817 VEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAV 876
Db    844 -----EDTTVVAVVESS-------GEQPASSSTSIP---------TELSK---------- 872

Qy    877 IDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEG 936
Db    873 -DDQVTEASGEETT----TAAATEASEETTTSA------VTEGS--GEETTVVAVVESSG 919

Qy    937 KD------------AHDDMVTSEVDFTSEAVTATE-TSEALRTEEVTEASGAEETTDMVS 983
Db    920 EEPASSSTSIPTELSKDDQVTEASGEETTTAAATEASSEETTTSAVTEGSGEETTTSAVT 979

Qy    984 AVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK----EESQVPATQ 1039
Db    980 EGS----GEETTTSAVP-----------EGENSTTEAPAFVTGSEIEIPSSEESSSTTH 1024

Qy   1040 -------TVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEH------LAQ 1086
Db   1025 DPSIPVITPKPSVSSTIENVMS-KTSSEEAAEKKIIGEHQTGKDDDAGKEDEDNMPAFVT 1083

Qy   1087 GSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLA 1146
Db   1084 ANPAGTSTTESAENVTSTGEED-----ENIKMAKELGKQFAAD---------------LA 1123

Qy   1147 DSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEP 1206
Db   1124 KLAAKDGVNLTETADAKDSGETAHVEDEQVSSTESSIGSEETTTTV--NKETTEEHHEAS 1181

Qy   1207 GRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVT 1266
Db   1182 G--------EEDDAPAFVTGAPTDSTTEASVSTTSA--ITDETTSVAADESTSTSAGEVQ 1231

Qy   1267 YDSEVMGVA--GCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAAPEH 1324
Db   1232 SSSAIIDSATVASEEQTSSEATSVIESSGE-EVTTTDENLVTSTVAQLEEGSGITAAESK 1290

Qy   1325 EGTYGKPVLTLDMPSSERGKALGSLGGSPS--LPDQDKAGCIEVQVQSLD----TTVTQT 1378
Db   1291 D----EDSVTTEATSQSTTVSESSDGSGESTVAPNDSETSTTESSQSTTDEGSGVTAAES 1346

Qy   1379 AEAVEKVIE--TVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTV---PLGPES 1433
Db   1347 KDEESSTTEAPAFVTSKTSGSEE---------DEEDSPDTHEFLTGIDETMFNKSLVPDT 1397

Qy   1434 QAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKV 1489
Db   1398 HREDLPNNVGFVPSSEPKPKNPDE---------EEEEEEDDGTKSDDYEDNVSKKI 1444
```

```
RESULT 10
O76891
ID   O76891      PRELIMINARY;      PRT;  5327 AA.
AC   O76891;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   EG:49E4.1 protein.
GN   FUTSCH OR EG:49E4.1 OR CG3064.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Papagiannakis G., Spanos L., Siden-Kiamos I., Louis C.;
RT   "Sequencing the distal X chromosome of Drosophila melanogaster.";
RL   Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Benos P.;
RL   Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL031128; CAA20006.1; -.
DR   FlyBase; FBgn0015390; futsch.
SQ   SEQUENCE   5327 AA;  575942 MW;  FEFEE23A118FF38A CRC64;

  Query Match             7.1%;  Score 574;  DB 5;  Length 5327;
  Best Local Similarity   21.6%;  Pred. No. 1.5e-14;
  Matches  415;  Conservative  285;  Mismatches  734;  Indels  484;  Gaps   82;

Qy        6 STEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPAD-------ADPATKLPQKNGQL 58
Db     2744 SVVESSKDDAEKSESRPESVIASGEPVPRESKSPLDSKDTSRPGSVVESVTAEDEKSEQQ 2803

Qy       59 SSVNGVAE--QGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMA----ANSTA 112
Db     2804 SRRESVAESVKADTKKDGKSQEASRPSSVDELL---KDDDEKQESRRQSITGSHKAMSTM 2860

Qy      113 VEDITKDGQEETSEIIEQIPASENNVEEMVQ-----------------PAESQANDV--- 152
Db     2861 GDESPMDKADKSKE-----PSRPESVAESIKHENTKDEESPLGSRRDSVAESIKSDITKG 2915

Qy      153 -----GFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEP---- 203
Db     2916 EKSPLPSKEVSRPESVVGSIKDEKAESRRESVAESVKPESSKDATSAPPSKEHSRPESVL 2975

Qy      204 -----------SVETAVGESASKESELKQSTE-----------KQEGTLKQEQSSTEIPL 241
Db     2976 GSLKDEGDKTTSRRVSVADSIKDEKSLLVSQEASRPESEAESLKDAAAPSQETSRPESVT 3035

Qy      242 QAESD-------------QAAEEEAKDEGEEKQEKEPTKSPESPSSPVNSETTSSFKKFF 288
Db     3036 ESVKDGKSPVASKEASRPASVAENAKDSADESKEQRPESLPQSKAGSIKDEKSP------ 3089

Qy      289 THGWAGWRKKTSFKKSKEDD-------------------------LETAEKRKEQ---- 318
Db     3090 ------LASKDEAEKSKEESRRESVAEQFPLVSKEVSRPASVAESVKDEAEKSKEESPLM 3143

Qy      319 ----------------EAEKVDEEEK-----EKTE-PASEEQEPAEDTDQARLSADYEKV 356
Db     3144 SKEASRPASVAGSVKDEAEKSKEESRRESVAEKSPLPSKEASRPASVAESVKDEADKSKE 3203

Qy      357 ELPLEDQVGDLEASSEEKCAP--LATEVFDE----KMEAHQEVVAEVH------------ 398
Db     3204 ESRRESGAEKSPLASKEASRPASVAESIKDEAEKSKEESRRESVAEKSPLPSKEASRPTS 3263

Qy      399 -----VSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELMKSR 453
Db     3264 VAKSVKDEAEKSKEESSRDSVAEKSPL--ASKEASRPASVAE--SVQDEAEKSKE--ESR 3317

Qy      454 EMCVSGGDHTQLTDLSPEEKTLPKH-PEGIVSEVEMLSSQERIK--VQGSPLKKLFSS-- 508
Db     3318 RESVA-----EKSPLAYKEASRPASVAESIKDEAEKSKEESRRESVAEKSPLASKEASRP 3372

Qy      509 -SGLKKLSGKKQKGKRGGGGDE--EPGEYQHIHTESPESADEQKGESSASSPEEPEETTC 565
Db     3373 TSVAESVKDEAEKSKEESSRDSVAEKSPLASKEASRPASVAESVQDEAEKSKEESRRESV 3432

Qy      566 LEKGPL---EAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRP--------- 613
Db     3433 AEKSPLASKEASRPASVAESVKDDAEKSKEESRRESVAEKSPLASKEASRPASVAESVKD 3492

Qy      614 -SESDKEE---ELEKVKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEAL 669
Db     3493 EAEKSKEESRRESVAEKSPLPSKEASRPTSVAESVKDEAEKSK-EESRRE---SVAEKSS 3548

Qy      670 ICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAV-PASTQE--QDQAQ 726
Db     3549 LASKEASRPASVAESVKDEA--EKSKEESRRESVAEKSPLASKEASRPASVAESVKDEAE 3606

Qy      727 GS----------SSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSV--- 772
Db     3607 KSKEVSRRESVAEKSPLPSKEASRPTSVA--ESVKD-EADKSKEESRRESGAEKSPLASM 3663

Qy      773 ---------EQLSTEIEPSREES---WVSIKKFIPGRRKKR----ADGKQEQA--TVEDS 814
Db     3664 EASRPTSVAESVKDETEKSKEESRRESVTEKSPLPSKEASRPTSVAESVKDEAEKSKEES 3723

Qy      815 GPVEINEDDP------NVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTL 868
Db     3724 RRESVAEKSPLASKESSRPASV-------AESIKDEAEGTKQ-----------ESRRESM 3765

Qy      869 VHTVSVAVIDGTR-AVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIA---- 923
Db     3766 PESGKAESIKGDQSSLASKETSRPDSVVESVKDETEKPEGSAIDKSQVASRPESVAVSAK 3825

Qy      924 -EETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMV 982
Db     3826 DEKSPL--HSRPE----------SVADKSPDA--SKEASRSLSVAE-TASSPIEEGP--- 3867

Qy      983 SAVSQLTDSPDTTEEA---------TPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQ 1034
Db     3868 RSIADLSLPLNLTGEAKGKLPTLSSPIDVAEGDFLEVKAESSPRPAVLSKPAEFSQPDTG 3927

Qy     1035 VPATQTVQRTGSKALEKVEEVEEDSE----VLASEKEKDVMPKGPVQEAGAEHLAQGSET 1090
Db     3928 HTASTPVDE-ASPVLEEIEVVEQHTTSGVGATGATAETDLL---DLTETKSETVTKQSET 3983

Qy     1091 G---------QATPESLE------VPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLT 1135
Db     3984 TLFETLTSKVESKVEVLESSVKQVEEKVQTSVKQAETTVTDSLEQLTK-----KSSEQLT 4038

Qy     1136 ------DSETNGSTPLADS--DTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKE 1187
Db     4039 EIKSVLDTNISNVTNLFSTAVETIEKKVQDVT-EKVIEKATEHVSEHVTTTGESSTETSQ 4097

Qy     1188 EPSTL--------------------------PNNVPA-------QEEHGEEPGRDV---- 1210
Db     4098 EKSSLDLGTFSELRETHITTVGSPEFTVTICERDEPVLHDIKEEDEEHRFSPPSDVDKAA 4157

Qy     1211 ---------LEPTQQELTAAVPVLAKTEVGQEGEVDWLDG-------EKVKEEQEVFVH 1254
Db     4158 IIPPQPMRPLSPREEEV-AKIVADVAKVLKSDKDITDIIPDFDERQLEEKLKSTADTEEE 4216

Qy     1255 SGPNS--QKAADVTYDSEVMGVAGCQEKESTEVQS--LSLEEGEMETDVEKEK------R 1304
Db     4217 SDKSTRDEKSLEISVKVEIE-----SEKSSPDQKSGPISIEEKDKIEQSEKAQLRQGILA 4271

Qy     1305 ETKPEQVSEEGE------QETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQ 1358
Db     4272 SSRPESVASQPESVPSPSQSAASHEHK-----------EVELSESHKAEKS-----SRPES 4316

Qy     1359 DKAGCIEVQVQSLDTTVTQTA--------------EAVEKVIETVVISETGESPECVGAHL 1405
Db     4317 -----VASQVSEKDMKTSRPASSTSQFSTKEGDEETTESLLHSLTTTETVETKQMEEKSS 4371
```

```
Qy     1406 LPAEKSSATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISASQRE- 1464
             :  :| |    |:  ::  || |   || :||:             |:|   :       :|:||
Db     4372 FESVSTSVTKS--TVLSSQSTVQLREESTSESL--------SSSLKVE-----DSSRRES 4416

Qy     1465 -RSEEEDKPDAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAVDQFARTE- 1522
             |   :|     :     || |:      :  |   :::  |   |:: :::  |||::  |  :  |
Db     4417 LSSLLAEKGGIATNTSLKEDTS-ASASQLEELLVQSEECSSESIVSEIQTSIAQKSNKEI 4475

Qy     1523 -TAPETHAYDSQTQVPACRLDSREPNRCWTKMKDAKMKHPVPQPREDLQVLTVLEAWA 1579
             | ||                  :: |:     :  ||  :|  |  :      ::::  ||::
Db     4476 KDARET-----------KVTSQFTTTTSSATKDDSLKETVAEFLATEKIVSAKEAFS 4521
```

RESULT 11
Q9FND5
ID   Q9FND5   PRELIMINARY;   PRT; 2910 AA.
AC   Q9FND5;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Similarity to heat shock protein.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=98069011; PubMed=9405937;
RA   Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA   Tabata S.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT   Sequence features of the regions of 1,044,062 bp covered by thirteen
RT   physically assigned P1 clones.";
RL   DNA Res. 4:291-300(1997).
DR   EMBL; AB006702; BAB11602.1; -.
DR   EMBL; AB017062; BAB11602.1; JOINED.
KW   Heat shock.
SQ   SEQUENCE   2910 AA; 325351 MW; A847EC3FE1427DF7 CRC64;

```
  Query Match             7.0%;  Score 564.5;  DB 10;  Length 2910;
  Best Local Similarity   21.6%;  Pred. No. 1.8e-14;
  Matches  404;  Conservative  275;  Mismatches  645;  Indels  549;  Gaps   84;

Qy       65 AEQG--DVHVQEENQEG--QEEEVVDEDVGQRES-----EDVREK---DRVEEMAANSTA 112
            |||    :| | :| :|  | | |        ||       ||:  |    | |||
Db      319 AEQSSREVTVDKEKEEDIIQNIEEVQESPSVMESPTIQGEDIESKASLDHEEEM------ 372

Qy      113 VEDITKDGQEETSEIIEQIPA--SENNVEE-------MVQPAESQANDVGFKKVFKFVGF 163
             : |||| :|:     :: :|      | |      || ||        ::   :    :|
Db      373 -DKITKDTEEQEHVLVRDVPVPQSETLVTEAKTAETFSVQEAEILKTNINESEAHSAIGG 431

Qy      164 K------------FTVKKDK-NEKSDTVQLLTVKKDEGEGAEASVGA-GDHQE------ 202
            :             :| ||  | |:||: : :  ||     ::      |:| |
Db      432 EEDGQETKENTEPSKDLKDDKEQEDSETVKTI-ISSDEVRSSDVQAEVFGEHTEPCSSEI 490

Qy      203 -------------------PSVETAV-GESASKESELKQSTEKQEGTLKQEQSSTEIPLQ 242
                               | :|||| |||   :      ::||:     ::|:    || :
Db      491 KDDRHGRDESIEVKAKETGPEIETAVDGESVHE----IETTERVLLEAEKEEDKEEIKID 546

Qy      243 AESDQAAEEEAKDEG-----EEKQ--EKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGW 295
            |     | |:|: |         || :       |  |   |||  |  |:|   :|   : |
Db      547 EEPSLNAIEKAETENVKIVIEEPEIVNNEETSVHESESLKENAEPVEAVKN--SDGTEQI 604

Qy      296 RKKTSFKKSKEDDL-----ETAEKRKEQEA-----EKVDE------------------- 325
             :: :  ::||:|:       |   :  |   |:||
Db      605 SREVTVDRAKEEDIAPKTEEIQERPSESKASLEPKEEVDHISNETEEHEHVLERDVQQCE 664

Qy      326 -------EEKEKTEPA---SEEQEPAE-DTDQARLSADYEKVELPLEDQVGDLEASSEEK 374
                   | || |:|:      |::|  | :|   |  :|  :    :|  |           |
Db      665 TIESEAVETKEDTQPSLDLKEDKETEEAETFKTVFSSDEVRSSAVQEEQFG-------EH 717

Qy      375 CAPLATEVFDE---KMEA----HQEVVAEVH------VSTVEKTEEEQGGGGEAEGGVVV 421
             | ::|: ||   | |:        || |  : :    |  |  |||: | | |  | |:|
Db      718 TEPCSSEIKDESHGKEESVEVKSQETVQDENTEDKHDVLKVPSTESEKYQGNEPE-TVLV 776

Qy      422 EGTG-------------------------------------------------------- 425
             ||
Db      777 SNTGSYEKSEKSPSDLVLNVDKEELNDEKINVDQVDGTQIMEEPIGLDSNGAEAEQIDQN 836

Qy      426 ---------------------ESLPPEKLAEPQEVPQEA----------EPAEELMKSR 453
                                  | :  ||  || | :|            :| ||:  :
Db      837 ITNETEEILVAKPVSLLDVKSVEQMQKPKLESPSEVSEETSKTVDEKIEEKPEEEVTLYQ 896

Qy      454 EMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKK--LFSSSGL 511
            | |  |           | :|:|:   || |  |:|    :||  :   :||:| | | |  :
Db      897 EGQVDGS-----YGLETKEETV-SVPESI--ELEEQPQEERSVIDPTPLQKPTLESPSEV 948

Qy      512 KKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKG-------ESSASSPEEPEETT 564
             : | |           :|  |    :    |  | | ||:         :  :| | | |: |
Db      949 LEESSKTV---------DEKIEEKTDSIELGEIAQEERSVTDLTPLQEESSQPNEQEKET 999

Qy      565 CLEK--GPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEEL 622
             |||       |  :  |   :: | | ||  |         | :  :    |::|:
Db     1000 KLEKHEPTNEEVKSDEVIEVLSASPSKELEGET--------VVEAENIENIKENEEEQAA 1051

Qy      623 EKVKSA-----TLSSTDSTV--SEMQDEVKTVGEEQKPEEPKRRV---------DTSVSW 666
             ||:: :    |: |  | :  || || || ||  |: |  |                | :
Db     1052 EKIQKSLETVQTVESPSSLLFSSEEQDHV-TVAEEIVDEKAKEEVPMLQIKNEDDATKIH 1110

Qy      667 EALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQ 726
             |  :       ::||   |  |         : ::  ||  |:          | ::|: : :
Db     1111 ETRV------EQARDIGPSLTE----ICSINQNQPEEQVKE---------ACSKEEQEKE 1151

Qy      727 GSSSPE---------PAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLST 777
             |:: |          :  :| |  :  ||     : |||   |  ||  ::  :|  :  |
Db     1152 ISTNSENIVNETYALHSVEAAEEETATNGESLDDVET----TKSVLLEVRKEEEEAEMKT 1207

Qy      778 EIEPS----REESWVSIKKFIPGRRKKRADGK---QEQATVEDSGPVE-INEDDPNVPAV 829
             : ||     :|  ::|: :            |  |:  |  :  || |    |  | |  |
Db     1208 DAEPRLDAIEKEELETVKTVV-------QDAKIVNNEETTAHESESLKGDNHQEKNAEPV 1260

Qy      830 VPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSK-----TLVHTVSVAVIDGTRAVT 884
                  : |:  |  :||         ::|:: |     |:: | ::    |   : |
Db     1261 EATQNLDDAEQISREVTVDTE-----READITEKIEKVQEGPTVIETPTIQGED-IESET 1314

Qy      885 SVEERSPSWISASVTEPLEHTAGEAMPPVE-----------------------EVTEKDI 921
             |:| :     |:  ||  ||    :| |                          ::| :
Db     1315 SLELKEEVDQSSKDTEEHEHVLERDIPQCETLKAEAVDTSTVEEAAILKTLETNISEPEA 1374

Qy      922 IAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTAT-------ETSEALRTEEVTEASG 974
             : || :  :   | |:|       | |::| |:         :|     :|    |: |
Db     1375 MHSETSLDLKVDKEQKEAE---TVKTVIFSNEVGTSDAQAEEFGEHTEPCSSEIKDESQG 1431

Qy      975 AEETTDM---------------------------------VSAVSQLTDSPDTTEEAT 999
             :||: ::                                         :| ||: |:: | ||
Db     1432 SEESVEVKSKETVQGESSEEKDVNMLDVQSGESEKYQENEPDISLVSK-TENGDKFEEIP 1490

Qy     1000 PVQE---------------VESGV---LDTEEEERQTQAILQAVADKVKEESQVPATQTV 1041
             | |                 ||| |    ||| ||  ::  | :  :  || |||   |    |
Db     1491 SVVEGAGLDETTHNQTLLDVESVVKQSLDTPSEEETSKTIDEKIEDKPKEE--VTLHQEG 1548

Qy     1042 QRTGSKALEKVEEV-------------EEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGS 1088
             : ||  |:  :|                 :::  ||:|:|: :    :  :|    :|
Db     1549 REEGSYGLDTKDEAVSVLESRELGEQPQQEELCLANEQENETKLQ---EEQVDKHEPTKE 1605

Qy     1089 ETG--QATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLA 1146
             |    | :|    |::  :|  |::  :          :   ||  ||:  ::|   |    :|
Db     1606 EVSNDQQSP----VEEISNEVIQVSSASL---------SEGPE-YETVVEAEKIGEEQVA 1651

Qy     1147 DSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGE-- 1204
            |     :  | :::   |  :::   :     |   |    :|            |   |  |:
Db     1652 DK-IQKSFETGEIVEAHSSLPSSSEEKEHETVSEKTDDEK---------VKDAEPIGDMR 1701
```

```
Qy     1205 EPGRDVLEPT---------QQELTAAAVPVLA-------KTEVGQEGEVDWLDGEKVKEE 1248
Db     1702 ERGLDIAETTHLSLPSVDQKEDVDEIHIPSVALPLDEQEKVTSTEKGETKSSEAEDDKPD 1761

Qy     1249 QEVFVHSGPN-SQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETK 1307
Db     1762 EHVDSSTSPMLSEKNDNETQTS----------KTSEDVCMQQEESGTLEVPKPEESKEDK 1811

Qy     1308 PEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSL------GGSP----SLPD 1357
Db     1812 SQEISETIEEIEATSDQ---------TLPIETSHTDNTLSSELVSEQDDQSPKKVEEIHE 1862

Qy     1358 QDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISETGE-----SPECVGAHLLPAEKSS 1412
Db     1863 EEPKEAHDVEATSERNLPVETSDA-DNTLSSQLVSETKEEHKLQAGEILPTEIIPRESSD 1921

Qy     1413 ATGGHWTLQHAEDTVPLGPESQAESIPIIVTPAPESTLHPDLQGEISAS-QRERSEEEDK 1471
Db     1922 EALVSMLASREDDKVALQEDNCADDV-------RETN---DIQEERSISVETEESVGETK 1971

Qy     1472 P----DAGPDADGKESTAIEKVLKAEPEILELESKSNKIVLNVIQTAV-----DQFARTE 1522
Db     1972 PKEHEDEIRDAHVETPTAPIILEENDSETLIAEAKKGNEEINETERTVALDHEEEFVNHE 2031

Qy     1523 TAPETHAYDSQTQ 1535
Db     2032 APKLEETKDEKSQ 2044
```

```
RESULT 12
Q9NFS3
ID   Q9NFS3      PRELIMINARY;      PRT; 16215 AA.
AC   Q9NFS3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   D-titin.
GN   SLS OR D-TITIN OR CG1915.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Zhang Y.Q., Broadie K.S.;
RT   "Characterization of Drosophila D-Titin gene.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AJ271740; CAB93524.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0003432; sls.
DR   InterPro; IPR002106; AAtRNA_ligaseII.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 50.
DR   Pfam; PF00018; SH3; 1.
DR   SMART; SM00408; IGc2; 15.
DR   SMART; SM00410; IG_like; 34.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS00179; AA_TRNA_LIGASE_II_1; UNKNOWN_1.
DR   PROSITE; PS50002; SH3; 1.
KW   Immunoglobulin domain; SH3 domain.
SQ   SEQUENCE   16215 AA;  1841509 MW;  242C8765E00F7603 CRC64;

  Query Match             6.9%;  Score 558;  DB 5;  Length 16215;
  Best Local Similarity   20.5%;  Pred. No. 2.3e-13;
  Matches  390;  Conservative  303;  Mismatches  720;  Indels  490;  Gaps   79;
```

```
Qy        7 TEQRSPEQPAG-----------SDTP-SELVLSGHGPAAEASGAAGDPADADPATKLPQK 54
Db    12422 TTKRMLRRPAGEGEIEIIEVVRDDQPEAEITIVEYEP---------EPVNQDEKPKEPKK 12472

Qy       55 NGQLSSVNGVAEQGDVH--VQ------------EENQEGQEEEVVDE-------DVGQRE 93
Db    12473 KTR------KVKKDDIHDYIQKLIELETPKTELEKYEKIEFEPIVKDKPLDSPIDVLDES 12526

Qy       94 SEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVG 153
Db    12527 PKEVQKKDKKSRTICKGQCVEEEAPEQPEIPVQILEVKPV-EVDVKEVITEDGKPVQEKT 12585

Qy      154 FKKVFKFVG------FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVET 207
Db    12586 TKRVLKKIGPEEQTTFKITM--IESEDNDSV---TVIVDEEPEIASPQSIEEHPEQSKEK 12640

Qy      208 AVGESAS-----KESEL--------------KQSTEKQEGTLKQEQSSTEIPLQAESDQAA 249
Db    12641 LAPKPKKTVRKVKKDDLSDYVKKLIEEEIPKVDLEKYE---KVEMPEKPVKLTV-SDSIP 12696

Qy      250 EEEAKDEGE------EKQEKEPTKSPESPSSPVNSETTSSFKKFFTHGWAGWRKKTSFKK 303
Db    12697 EEPKPDKSQPISVLPDTTKPKKTKTPKTPKTEDTDQQVPDEPTETTVDTTDIPELTPTQT 12756

Qy      304 SKEDDLETAEKRKEQEAEKVDEEE-----------KEKTEPASEEQEPAEDTDQARLSADY 353
Db    12757 AQPEDTATAQITPSAQEEKSTQDDTKDTIQKTVKHKKTKP--DTQKSVETSELPEVHKDY 12814

Qy      354 ------------EKVELPLEDQVGDLEASSEEKCAPLATEVFDE---------------- 385
Db    12815 QISIIHEELVEEEQPEKILEVRVID-EVAEVEESQPIVEEVEDEEPQPATEETVEDVTKP 12873

Qy      386 ----------KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAE 435
Db    12874 KSKKKKVVKKKTDDHDELIKKMLEQEIEKTELEKYEKIEFDVPKKLKPEFAALEPIKIER 12933

Qy      436 PQEVP---------------------QEAEPAEEL---------MKSREMCVSGGDHTQLT 466
Db    12934 KEQKPTKVTILDATDVPKTVKLKPSKRKEKPAEELTVQLPKFRLKARMVLVE-------- 12985

Qy      467 DLSPEEKTLPKHPE-GIVSEVEMLS-----SQERIKVQGSPLKKLFSSSGLKKLSGKKQK 520
Db    12986 --YPPAPLIPKTTDIGAIKDNGELSRNIGEAEEILKFKPHKTKKI-----------KKIK 13032

Qy      521 GKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLEKGPLEAPQDGEAE 580
Db    13033 -------------------DDLEKVELEKYEKYISSEEEPEEKTPYKK-PEKAPKPEEKQ 13072

Qy      581 EGT---TSDGEKK-REGITPWASFKKMVTPKKRVRRPSESDKE-------EELEKVKSAT 629
Db    13073 EDVKLKLGKGKKKPKEEEAP-----ENVTLKNIPQKPQEVEEEVELKQKPKEVEIVEEQT 13127

Qy      630 LSSTDSTV-------SE------MQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSK 676
Db    13128 KKPKDGEFVVEPFEPSEFDRPEYVPDELEQIEHPEIPEKVKKPSKT-------------K 13174

Qy      677 KRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGS 736
Db    13175 YKPKDKSKSEPETIVSEIVAGVPKEEEAIPEQDVKFRKPERDAPEETDSEIKLRPVPQAS 13234

Qy      737 PSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVE-------QLSTEIEPSREESWVS 789
Db    13235 KDEN------PDEQALVTP--KAEEPIPQEIEDKAIDDEKKPKKSKPKKVQPKEQE---- 13282

Qy      790 IKKFIPGRRKKRADGKQEQATVEDSGPVEINE-DDPNV----PAVVPLSEYNAVEREKME 844
Db    13283 IAKEEP--EEFEVSVKEEEALVDK--PIEIEKPKDVKVKEKKPKEAPVSEVVVIEEEPKP 13338

Qy      845 AQGNTELP---QLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTS--VEER---------- 889
Db    13339 EEVPEEIPVEYKITTTVLEPEDAPKE--HQVKVIDFDERQETTEEVIEEKVVTRKKKPKP 13396

Qy      890 -SPSWISASVTEPLEH------TAGEAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDD 942
```

```
                  |     : :  | |  |          :  |      | : | |   : |        |        : | | |  | |    :  : |
Db  13397 QQPEEFEVTLKEPKEEQIQPDVVSAEISLPIEEPEQKPEQYEVELKITQTTPE--EPNDV 13454

Qy    943 MVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQ--LTDSPDTTE-EAT 999
                :   :     : :  |    :   :   :    |      |      | | |     |        :    : :  |  :  |  :  :
Db  13455 QIAVKEKVKTKPVKKVKEDKIVVVEAEEEKQPVEETIVEVEKQEEKKKSEKPKSYEFKIS 13514

Qy   1000 PVQEVESGVLDTEEE--ERQTQAILQAVADK-------VKE-------------ESQVPA 1037
                 | :|    : :   | |   |     :  :  :  | | : |            : | |                 |  :  |
Db  13515 ETQSIEEKPIEVAEEAPEETPKVVEKKVAEKFDSYEFTLKETDEEKVITVDDQPEEEAP- 13573

Qy   1038 TQTVQRTGSKALEKVEE---------VEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGS 1088
                 :  |  :       |   |  | |              | | |   |     |  : : : |  .  | |       : | |           |
Db  13574 VEVVFKKKPKEPEAVEAEFVMTEPKIVEETSVETAIKQKKTKKPKKDEEEAQLAIKVVES 13633

Qy   1089 ETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLT----DSETNGSTP 1144
                 |      |          |  | |              : : : :      | :  : |  |    :    |       : | |           |
Db  13634 EAPVAEEVFSEAPE----------SKIVE-----EEVIAEEKPKEFTIRVSESEPKPEEP 13678

Qy   1145 LADSDTADGTQQDET--------IDSQDSKATAAVRQ----------SQVTEEEAATAQK 1186
                  :       :      |             |    : : | |     |   :                   : | | :   | |      :
Db  13679 SVEQFTVKKRKPSVTFADEPATEIVIKESKPAEVVTEDAHIKTKKPKKKVTDVEAEELKI 13738

Qy   1187 EEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKT-EVGQEGEVDWLDGEKV 1245
                 :       : |    : |    | |    | |      : | :    | : :        |     |     | |   : : |  :                |  :
Db  13739 KITEEVPQEIPILEEVSEE---EVITETKK----TAPVVEEKTYKIGIKETEPEKPAEAI 13791

Qy   1246 KEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRE 1305
                   | | :  |              :  |    :   :  |      :                  |          |    |    :  |  :  | :    :  | :
Db  13792 VEEEEPVVTEPIEEAPKPEVFEEHKVRVI---------EETPRELVEEVIEEEVKVIRRK 13842

Qy   1306 TKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIE 1365
                  : :  | |  |           :     |  | |                |  :    |   :     |      |   : : : |     :  :
Db  13843 KPKPEIKEEPEAEVTV-----STPKPV---------EEVEATSSIAVIPEQPTEEEAADLK 13889

Qy   1366 VQVQSLDTTVTQTAEAVEKVIETVVISE--------------TGESPECVGAHLLPAEKS 1411
                 :   :     :          |       |  | : : :   |     :  :   |                       |  : |    :          | |   | :
Db  13890 ITIIEEE---TPPQELVQEIEEIEIVEEPKAPEEQPTDFTFATKDSEKKPTVEELPEEQV 13946

Qy   1412 SATGGHWTLQHAEDTVPLG---PESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEE 1468
                              | : |     :       | :          |  : | |   :       | |     |     |          :   : | :   | : :         | |
Db  13947 -------TIQKKKKKAPVPEVVEEPEAEFLVKPKTPVQEVT----EEAKITKSKKPVKEE 13995

Qy   1469 EDKPDAGPDADGKESTAIEKVLKAEPEILEL--------ESKSNKIVLNVIQTAVDQFAR 1520
                 |                   |       :      |   :   :     | | | : | :                 | |      :    |  : |    : :     :
Db  13996 E--------AAAELKVTITEEIPTEPEVQEIIEEIEEIEEEKPAEYVIEVKESQPEAVED 14047

Qy   1521 TET--------APETHAYDSQ-TQVPACRLDSREPNRCWTKMKDAKM------------- 1558
                  |                        | |          : : :  |     |        :    :     :          |   |     | :
Db  14048 KEVSLPKKKPKAPIVEEPEAEITLKPKVKSEEVQEEAKIVKKKPKKIDEVAVADELTVKV 14107

Qy   1559 -KHPVPQPREDLQVLTVLEAWAQPRKCLPR-------LQLKAP 1593
                  :      | | : |     :  : | :       |       : | : :       |               : : | |   |
Db  14108 EEEVVPEPIVEEEVIEEFEIKKKPKEPEPEDIVDAAIVKLKKP 14150
```

RESULT 13

```
Q9I7U4
ID   Q9I7U4      PRELIMINARY;      PRT;  6815 AA.
AC   Q9I7U4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   CG18242 protein.
GN   CG18242.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC   Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BERKELEY;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
CC   -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR   EMBL; AE003473; AAG22226.1; -.
DR   HSSP; P56276; 1TLK.
DR   FlyBase; FBgn0035301; CG18242.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001452; SH3.
DR   Pfam; PF00041; fn3; 5.
DR   Pfam; PF00047; ig; 11.
DR   Pfam; PF00018; SH3; 1.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 2.
DR   SMART; SM00408; IGc2; 5.
DR   SMART; SM00410; IG_like; 6.
DR   SMART; SM00326; SH3; 1.
DR   PROSITE; PS50002; SH3; 1.
KW   Immunoglobulin domain; Repeat; SH3 domain.
SQ   SEQUENCE   6815 AA;  779559 MW;  A4E244001A4EBA01 CRC64;

  Query Match             6.9%;  Score 557.5;  DB 5;  Length 6815;
  Best Local Similarity   20.5%;  Pred. No. 8.9e-14;
  Matches  393;  Conservative  304;  Mismatches  716;  Indels  503;  Gaps   80;

Qy      7 TEQRSPEQPAG-----------SDTP-SELVLSGHGPAAEASGAAGDPADADPATKLPQK 54
              | :|    :|||              | | :|: :   :   |              :|  :  |    | |:|
Db   2989 TTKRMLRRPAGEGEIEIIEVVRDDQPEAEITIVEYEP---------EPVNQDEKPKEPKK 3039

Qy     55 NGQLSSVNGVAEQGDVH--VQ------------EENQEGQEEEVVDE-------DVGQRE 93
               :               : :  |:|   :|             |:  : :   |  :|  :             ||
Db   3040 KTR------KVKKDDIHDYIQKLIELETPKTELEKYEKIEFEPIVKDKPLDSPIDVLDES 3093
```

```
Qy     94 SEDVREKDR-------------VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEE 140
Db   3094 PKEVQKKDKKSRSTKVPNEETPVQEQYAKVNVVEEEAPEQPEIPVQILEVKPV-EVDVKE 3152

Qy    141 MVQPAESQANDVGFKKVFKFVG------FKFTVKKDKNEKSDTVQLLTVKKDEGEGAEAS 194
Db   3153 VITEDGKPVQEKTTKRVLKKIGPEEQTTFKITM--IESEDNDSV---TVIVDEEPEIASP 3207

Qy    195 VGAGDHQEPSVETAVGESAS-----KESEL-------------KQSTEKQEGTLKQEQSS 236
Db   3208 QSIEEHPEQSKEKLAPKPKKTVRKVKKDDLSDYVKKLIEEEIPKVDLEKYE---KVEMPE 3264

Qy    237 TEIPLQAESDQAAEEEAKDEGE------EKQEKEPTKSPESPSSPVNSETTSSFKKFFTH 290
Db   3265 KPVKLTV-SDSIPEEPKPDKSQPISVLPDTTKPKKTKTPKTPKTEDTDQQVPDEPTETTV 3323

Qy    291 GWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEE----------KEKTEPASEEQEP 340
Db   3324 DTTDIPELTPTQTAQPEDTATAQITPSAQEEKSTQDDTKDTIQKTVKHKKTKP--DTQKS 3381

Qy    341 AEDTDQARLSADY------------EKVELPLEDQVGDLEASSEEKCAPLATEVFDE--- 385
Db   3382 VETSELPEVHKDYQISIIHEELVEEEQPEKILEVRVID-EVAEVEESQPIVEEVEDEEPQ 3440

Qy    386 -----------------------KMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVE 422
Db   3441 PATEETVEDVTKPKSKKKKVVKKKTDDHDELIKKMLEQEIEKTELEKYEKIEFDVPKKLK 3500

Qy    423 GTGESLPPEKLAEPQEVP---------------------QEAEPAEEL--------MKSR 453
Db   3501 PEFAALEPIKIERKEQKPTKVTILDATDVPKTVKLKPSKRKEKPAEELTVQLPKFRLKAR 3560

Qy    454 EMCVSGGDHTQLTDLSPEEKTLPKHPE-GIVSEVEMLS-----SQERIKVQGSPLKKLFS 507
Db   3561 MVLVE----------YPPAPLIPKTTDIGAIKDNGELSRNIEEAEEILKFKPHKTKKI-- 3608

Qy    508 SSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCLE 567
Db   3609 ---------KKIK------------------DDLEKVELEKYEKYISSEEEPEEKTPYK 3640

Qy    568 KGPLEAPQDGEAEEGT---TSDGEKK-REGITPWASFKKMVTPKKRVRRPSESDKE---- 619
Db   3641 K-PEKAPKPEEKQEDVKLKLGKGKKKPKEEEAP-----ENVTLKNIPQKPQEVEEEVELK 3694

Qy    620 ---EELEKVKSATLSSTDSTV-------SE------MQDEVKTVGEEQKPEEPKRRVDTS 663
Db   3695 QKPKEVEIVEEQTKKPKDGEFVVEPFEPSEFDRPEYVPDELEQIEHPEIPEKVKKPSKT- 3753

Qy    664 VSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQD 723
Db   3754 ------------KYKPKDKSKSEPETIVSEIVAGVPKEEEAIPEQDVKFRKPERDAPEET 3801

Qy    724 QAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVE-------QLS 776
Db   3802 DSEIKLRPVPQASKDEN------PDEQALVTP--KAEEPIPQEIEDKAIDDEKKPKKSKP 3853

Qy    777 TEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGPVEINE-DDPNV----PAVVP 831
Db   3854 KKVQPKEQE----IAKEEP--EEFEVSVKEEEALVDK--PIEIEKPKDVKVKEKKPKEAP 3905

Qy    832 LSEYNAVEREKMEAQGNTELP---QLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTS--V 886
Db   3906 VSEVVVIEEEPKPEEVPEEIPVEYKITTTVLEPEDAPKE--HQVKVIDFDERQETTEEVI 3963

Qy    887 EER-----------SPSWISASVTEPLEH------TAGEAMPPVEEVTEKDIIAEETPVL 929
Db   3964 EEKVVTRKKKPKPQQPEEFEVTLKEPKEEQIQPDVVSAEISLPIEEPEQKPEQYEVELKI 4023

Qy    930 TQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEASGAEETTDMVSAVSQ-- 987
Db   4024 TQTTPE--EPNDVQIAVKEKVKTKPVKKVKEDKIVVVEAEEEKQPVEETIVEVEKQEEKK 4081

Qy    988 LTDSPDTTE-EATPVQEVESGVLDTEEE--ERQTQAILQAVADK-------VKE------ 1031
Db   4082 KSEKPKSYEFKISETQSIEEKPIEVAEEAPEETPKVVEKKVAEKFDSYEFTLKETDEEKV 4141

Qy   1032 -------ESQVPATQTVQRTGSKALEKVEE---------VEEDSEVLASEKEKDVMPKGP 1075
Db   4142 ITVDDQPEEEAP-VEVVFKKKPKEPEAVEAEFVMTEPKIVEETSVETAIKQKKTKKPKKD 4200

Qy   1076 VQEAGAEHLAQGSETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLT 1135
Db   4201 EEEAQLAIKVVESEAPVAEEVFSEAPE----------SKIVE-----EEVIAEEKPKEFT 4245

Qy   1136 ----DSETNGSTPLADSDTADGTQQDET--------IDSQDSKATAAVRQ---------- 1173
Db   4246 IRVSESEPKPEEPSVEQFTVKKRKPSVTFADEPATEIVIKESKPAEVVTEDAHIKTKKPK 4305

Qy   1174 SQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDVLEPTQQELTAAAVPVLAKT-EVG 1232
Db   4306 KKVTDVEAEELKIKITEEVPQEIPILEEVSEE---EVITETKK----TAPVVEEKTYKIG 4358

Qy   1233 QEGEVDWLDGEKVKEEQEVFVHSGPNSQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEE 1292
Db   4359 IKETEPEKPAEAIVEEEEPVVTEPIEEAPKPEVFEEHKVRVI---------EETPRELVE 4409

Qy   1293 GEMETDVEKEKRETKPEQVSEEGEQETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGS 1352
Db   4410 EVIEEEVKVIRRKKPKPEIKEEPEAEVTV-----STPKPV--------EEVEATSSIAVI 4456

Qy   1353 PSLPDQDKAGCIEVQVQSLDTTVTQTAEAVEKVIETVVISE--------------TGESP 1398
Db   4457 PEQPTEEEAADLKITIIEEE---TPPQELVQEIEEIEIVEEPKAPEEQPTDFTFATKDSE 4513

Qy   1399 ECVGAHLLPAEKSSATGGHWTLQHAEDTVPLG---PESQAESIPIIVTPAPESTLHPDLQ 1455
Db   4514 KKPTVEELPEEQV-------TIQKKKKKAPVPEVVEEPEAEFLVKPKTPVQEVT----EE 4562

Qy   1456 GEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEPEILEL---------ESKSNKIV 1507
Db   4563 AKITKSKKPVKEEE--------AAAELKVTITEEIPTEPEVQEIIEEIEEIEEEKPAEYV 4614

Qy   1508 LNVIQTAVDQFARTET--------APETHAYDSQ-TQVPACRLDSREPNRCWTKMKDAKM 1558
Db   4615 IEVKESQPEAVEDKEVSLPKKKPKAPIVEEPEAEITLKPKVKSEEVQEEAKIVKKKPKKI 4674

Qy   1559 --------------KHPVPQPREDLQVLTVLEAWAQPRKCLPR-------LQLKAP 1593
Db   4675 DEVAVADELTVKVEEEVVPEPIVEEEVIEEFEIKKKPKEPEPEDIVDAAIVKLKKP 4730
```

RESULT 14
Q9LH98

```
ID   Q9LH98      PRELIMINARY;      PRT;  2081 AA.
AC   Q9LH98;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Genomic DNA, chromosome 3, BAC clone: T19N8.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RA   Kaneko T., Kato T., Sato S., Nakamura Y., Asamizu E., Tabata S.;
RL   Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=COLUMBIA;
RX   MEDLINE=20363099; PubMed=10907853;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. II.
RT   Sequence features of the regions of 4,251,695 bp covered by ninety P1,
```

```
RT   TAC and BAC clones.";
RL   DNA Res. 7:217-221(2000).
DR   EMBL; AP002057; BAB03174.1; -.
SQ   SEQUENCE   2081 AA;  232851 MW;  D3603E1F85EFFF29 CRC64;

  Query Match             6.4%;  Score 517;  DB 10;  Length 2081;
  Best Local Similarity   19.8%;  Pred. No. 9.4e-13;
  Matches  324;  Conservative  293;  Mismatches  721;  Indels  302;  Gaps   55;

Qy        5 SSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPADADPATKLPQKNGQ------L 58
Db      191 ASSEESSSTEHNNVTTGSNMV-ETNGENSESTQEKGDGVEGSNGGDVSMENLQGNKVEDL 249

Qy       59 SSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITK 118
Db      250 KEGNNVVENGE--TKENNGENVESNNEKEVEGQGES-------------IGDSAIE---- 290

Qy      119 DGQEETSEIIEQIPASENNVEEMVQP-AESQANDVGFKKVFKFVGFKFTVKKDKN----- 172
Db      291 KNLESKEDVKSEVEAAKNDGSSMTENLGEAQGNN----------GVS-TIDNEKEVEGQG 339

Qy      173 ---EKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESELKQSTEKQEGT 229
Db      340 ESIEDSDIEKNLESKEDVKSEVEAAKNAGSSMTGKLEEAQRNNGVSTNETMNSENKGSG- 398

Qy      230 LKQEQSSTEIPLQAESDQAAEEEAKDE----------GEEKQEK---EPTKSPESPSSPV 276
Db      399 ---ESTNDKMVNATTNDEDHKKENKEETHENNGESVKGENLENKAGNEESMKGENLENKV 455

Qy      277 NSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQ--------------EAEK 322
Db      456 GNEELK--------GNASVEAKTNNESSKEEKREESQRSNEVYMNKETTKGENVNIQGES 507

Qy      323 VDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEV 382
Db      508 IGDSTKDNSLENKEDVKPKVDANESDGNSTKERHQ---EAQVNN-GVSTEDKNLDNIGAD 563

Qy      383 FDEKMEAHQEVVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQE 442
Db      564 EQKKNDKSVEVTTNDGDHTKEKREETQGNNGES---VKNENLENKEDKKELKDDESVGAK 620

Qy      443 AEPAEELMKSREMCVSGGD---HTQLTDLSPEEKTLPKHPE---GIVSEVEMLSSQERIK 496
Db      621 TNNETSLEEKREQTQKGHDNSINSKIVDNKGGNADSNKEKEVHVGDSTNDNNMESKEDTK 680

Qy      497 VQGSPLKKLFSSSGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASS 556
Db      681 ----------SEVEVKKNDGSSEKGEEGKENNKDSMEDKKL--ENKESQTDSKDDKSVDD 728

Qy      557 PEEPEETTCLEKGPLEAPQDGEAEEGTTSDGE-KKREGITPWASFKKMVTPKKRVR---- 611
Db      729 KQEEAQI-----------YGGESKDDKSVEAKGKKKES----KENKKTKTNENRVRNKEE 773

Qy      612 ----RPSESDKEEELEKVKSATLSSTDS------TVSEMQDEVKTVGEEQKPEEPKRRVD 661
Db      774 NVQGNKKESEKVEKGEKKESKDAKSVETKDNKKLSSTENRDEAK----ERSGEDNKEDKE 829

Qy      662 TSVSWEALICVGSSKKRARKASSSDDEGGPRTLGGDSH---------------------- 699
Db      830 ESKDYQSV-----------EAKEKNENGGVDTNVGNKEDSKDLKDDRSVEVKANKEESMK 878

Qy      700 -RAEEASKDKEAGTDAVP--ASTQEQDQAQGSSSPEPAGSPSEGEGVSTWESFKRLVTPR 756
Db      879 KKREEVQRNDKSSTKEVRDFANNMDIDVQKGSGESVKYKKDEKKEG--NKEENKDTINTS 936

Qy      757 KKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRRKKRADGKQEQATVEDSGP 816
Db      937 SKQKGKDKKKKKK---ESKNSNMKKKEEDK----KEYVNNELKKQEDNKKETTKSENSKL 989

Qy      817 VEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLGAVYVSEELSKTLVHTVSVAV 876
Db      990 KEENKDNKEKKKESEDSASKNREKKEYEEKKSKTK-----------EEAKKEKKKSQ---- 1034

Qy      877 IDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVEEVTEKDIIAEETPVLTQTLPEG 936
Db     1035 -DKKREEKDSEERK------SKKEKEESRDLKAKKKEEETKEK----KESENHKSKKKED 1083

Qy      937 KDAHDDMVTSEVDFTSEAVTATETSEALRTEE---------------VTEASGAEETTDM 981
Db     1084 KKEHEDNKSMKKEEDKKEKKKHEESKSRKKEEDKKDMEKLEDQNSNKKKEDKNEKKKSQH 1143

Qy      982 VSAVSQLTDSPD--TTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVKEESQVPATQ 1039
Db     1144 VKLVKKESDKKEKKENEEKSETKEIESSKSQKNEVDKKEK--------KSSKDQQKKKEK 1195

Qy     1040 TVQRTGSKALEKVEE-------VEEDSEVLASEKEKDVM---PKGPVQEAGAEHLAQGSE 1089
Db     1196 EMKESEEKKLKKNEEDRKKQTSVEENKKQKETKKEKNKPKDDKKNTTKQSGGKKESMESE 1255

Qy     1090 TGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLADSD 1149
Db     1256 SKEA--ENQQKSQATTQADSDESKNEILMQ---ADSQADSHSDSQADSDESKNEILMQAD 1310

Qy     1150 TADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQ----EEHGEE 1205
Db     1311 SQATTQRNNEEDRKKQTSVAENKKQKETKEE-----KNKPKDDKKNTTKQSGGKKESMES 1365

Qy     1206 PGRDVLEPTQQELTAAAVPVLAKTEVGQEGEV-------DWLDGEKVKEEQEVFVHSGPN 1258
Db     1366 ESKEAENQQKSQATTQADSDESKNEILMQADSQADSHSDSQADSDESKNEILMQADSQAT 1425

Qy     1259 SQKAADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGE-- 1316
Db     1426 TQRNNEEDRKKQTSVAENKKQKETKEEKNKPKDDKKNTTEQSGGKKESM-ESESKEAENQ 1484

Qy     1317 ---QETAAPEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKA-GCIEVQVQSLD 1372
Db     1485 QKSQATTQGESDESKNEILMQADSQADTHANSQG---------DSDESKNEILMQADSQA 1535

Qy     1373 TTVTQTAEAVEKVI---ETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPL 1429
Db     1536 DSQTDSDESKNEILMQADSQADSQT-DSDESKNEILMQADSQAKIG-----ESLEDNKVK 1589

Qy     1430 GPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKV 1489
Db     1590 GKEDNGDEV------GKENSKTIEVKGRHEESKDGKTNENGGKEVSTEEGSKDSNIVERN 1643

Qy     1490 LKAEPEILELESKSNKIVLN 1509
Db     1644 GGKEDSIKEGSEDGKTVEIN 1663

RESULT 15
P91257
ID   P91257     PRELIMINARY;      PRT;  3484 AA.
AC   P91257;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   Hypothetical 385.7 kDa protein.
GN   F12F3.3.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
```

```
RA   Fulton B., Wohldmann P.;
RT   "The sequence of C. elegans cosmid F12F3.";
RL   Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BRISTOL N2;
RA   Waterston R.;
RT   "Direct Submission.";
RL   Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U80022; AAC25885.2; -.
DR   InterPro; IPR001064; Crystallin.
DR   InterPro; IPR003962; FnIII_repeat.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR003599; Ig.
DR   InterPro; IPR003598; Ig_c2.
DR   InterPro; IPR003600; Ig_like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR000130; Zn_MTpeptdse.
DR   Pfam; PF00041; fn3; 7.
DR   Pfam; PF00047; ig; 10.
DR   PRINTS; PR00014; FNTYPEIII.
DR   SMART; SM00060; FN3; 8.
DR   SMART; SM00409; IG; 12.
DR   SMART; SM00408; IGc2; 8.
DR   SMART; SM00410; IG_like; 1.
DR   PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; UNKNOWN_1.
DR   PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW   Hypothetical protein; Immunoglobulin domain; Repeat.
SQ   SEQUENCE   3484 AA;  385746 MW;  21DD6F6893020F4F CRC64;

  Query Match             6.2%;  Score 501.5;  DB 5;  Length 3484;
  Best Local Similarity   20.3%;  Pred. No. 7e-12;
  Matches  333;  Conservative  285;  Mismatches  560;  Indels  461;  Gaps   69;

Qy     80 QEEEVVDEDVGQRESEDVREKDRVEEMAANSTAVEDITKDGQEETSEIIEQIPASENNV- 138
Db    269 RDGKLIDDGSGDR--MEVRHED---EMRKHWLILKDICKD--EEAEYACQAI-----NVA 316

Qy    139 -------EEMVQPAESQANDVGFKKVFKFVGFKFTVKKDKN-EKSDTVQLLTVKKDEGEG 190
Db    317 GEAWCFSDVVVHMSEESRDD------------------DKSVDEVDDSTVLEEKKDDGD- 357

Qy    191 AEASVGAGDHQEPSV-ETAVGESASKESELKQSTEKQEGTLKQ--EQSSTEIPLQAESDQ 247
Db    358 --------DKSKPKTKKKIIKKKETPESEQVTAAEPEQQKISEVDVQSVAETEVGAKKKP 409

Qy    248 AAEE-----EAKDEGEEKQEKEPTKSPESPSS---PVNSETTSSFKKFFTHGWAGWRKKT 299
Db    410 DAEKPTDLSKAKKDSKSKKSDEPEASTEEKSTTEKPTNDKTS---------------KKS 454

Qy    300 SFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPASEEQEPAEDTDQARLSADYEKVELP 359
Db    455 AEKKTVKPKKEVT--GKPLEAKKPVEDKKDASQPSS-SKESSPPTDGKKKKQIPKALFIP 511

Qy    360 --LEDQVGD----------------LEASSEEK---CAPLATEVFDEKMEAHQEVVAEVH 398
Db    512 DEISSRFGDPSTMHSETNITTTIRGREGSADAKTPLVEPLSASV-SMKVESAKE---KAE 567

Qy    399 VSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQ-EVPQEAEPAEELMKSREMCV 457
Db    568 FSFKRRSETPDDKSRKKEG----------LPPAKKSEKKDEVTAEKQSTEALIESKKKEV 617

Qy    458 SGGDHTQLTDLSPEEKT------LPKHPEG-----IVSEVEMLSSQERIKVQGSPLKKLF 506
Db    618 ---DESKISEQQPSDKNKSEVVGVPEKAAGPETKKDVSEIEEV-----------PKKKTI 663

Qy    507 SSSGLKKLSGKKQKGKRGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL 566
Db    664 KKKTEKSDSSISQKSNVLKPADDDKSKSDDVTDKSKKTTEDQTKVATDSKLEKAADTT-- 721

Qy    567 EKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEKVK 626
Db    722 KQIETETVVDDKSKKKVL---KKKTEKSDSFISQKSETPPVVEPTKPAESEAQKIAEVNK 778

Qy    627 SATLSSTDSTVS-EMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKASSS 685
Db    779 AKKQKEVDDNLKREAEVAAKKIADEKLKIEAEANIKKTAEVEA----------AKKQKEK 828

Qy    686 DDE-------GGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPS 738
Db    829 DEQLKLETEVVSKKSAAEKLELEKQAQIKKAAEADAVKKQKELNEKNKLEAAKKSAADKL 888

Qy    739 EGEGVSTWESFK---RLVTPRKKSKSKLEEK-----AEDSSVEQLSTEIEPSREESWVSI 790
Db    889 KLEEESAAKSKKVSEESVKFGEEKKTKAGEKTVQVESEPTSKKTIDTKDVGATEPADETP 948

Qy    791 KKFIPGRRKKRADGKQEQATVEDSGPV--EINEDDPNVPAVVPLSEYNAV-EREKMEAQG 847
Db    949 KKKIIKKKTEKSDSSISQKSATDSEKVSKQKEQDEPTKPAV---SETQMVTEADKSKKQK 1005

Qy    848 NTELPQLLGAVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAG 907
Db   1006 ETD----------EKLK----------LDAEIAAKTKQEAD-----------EKSKL 1031

Qy    908 EAMPPVEEVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTE 967
Db   1032 DAQEKIKKVSEDDAARKE-----------KELND-----KLKLESEIATKKASADKLKLE 1075

Qy    968 EVTEASGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTE-----------EEER 1016
Db   1076 EQAQAKKAAE----VEAAKKQKEKDEQLK-----------LDTEAASKKAAAEKLELEK 1119

Qy   1017 QTQAILQAVAD------------------------KVKEESQVPATQTVQ-------RT 1044
Db   1120 QAQIKKAAGADAVKKQKELDEKNKLEANKKSAAGKLKIEEESAAKSKQTVEEQAKLDAQT 1179

Qy   1045 GSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQ-------EAGAEHLAQGSETGQATPES 1097
Db   1180 KAKTAEKQTKLEKDEKSTKESESKETVDEKPKKKVLKKKTEKSDSSISQKSETSKTVVES 1239

Qy   1098 LEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSET--LTDS-----ETNGSTPLADSDT 1150
Db   1240 ----------------------------AGPSESETQKVADAARKQKETDEKQKLEAEIT 1271

Qy   1151 ADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGEEPGRDV 1210
Db   1272 AKKSADEK------SKLEAESKLKKAAEVEAAKKQKEKDEQLKLDT---EAASKKAAAEK 1322

Qy   1211 LEPTQQELTAAAVPVLA---------KTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQK 1261
Db   1323 LELEKQSHIKKAAEVDAVKKQKELEEKQRLESEAATKKADAEKLKLEEQ--------KKK 1374

Qy   1262 AADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETAA 1321
Db   1375 AAEIA----LIEIQKEQEKLAQE-QSRLEDEAKKSAEKQKLESETKSKQTEE-------A 1422

Qy   1322 PEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQ---- 1377
Db   1423 PKE----------------------------SVDEKPKKKVLKKKTEKSDSSISQKSKS 1453

Qy   1378 ---TAEAVE------KVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVP 1428
Db   1454 AKSTVDAAETLESDFNLVEKKTVQKVEQSPD---------ESTSATIKRDPAQKTEEISK 1504

Qy   1429 LGPESQAESIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDA-GPDA--------- 1478
Db   1505 QDDGDEKKTTTDGKPPKPEDS-------EATPKKRVVKKKTQKSDSVASDASLADVSKLS 1557

Qy   1479 DGKESTAIEKVLK------------------AEPEILELES-KSNKIVLNVIQTAVDQFA 1519
Db   1558 DDVEEKPKKKVLKKKTEKSDSVISETSSVDTIKPESVEIPTEKAEQMILH------NRFS 1611

Qy   1520 RTETAPET---HAYDSQTQ 1535
Db   1612 -TDSAVESEPKNAHKDDTE 1629
```

Search completed: December 13, 2002, 00:16:01
Job time : 166 secs

GenCore version 5.1.3


```
OM protein - nucleic search, using frame_plus_p2n model

Run on:          December 13, 2002, 00:17:16 ; Search time 8528 Seconds
                                          (without alignments)
                                          5446.538 Million cell updates/sec

Title:           US-09-902-432-4
Perfect score:   8073
Sequence:        1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:   BLOSUM62
                 Xgapop 10.0 , Xgapext  0.5
                 Ygapop 10.0 , Ygapext  0.5
                 Fgapop  6.0 , Fgapext  7.0
                 Delop   6.0 , Delext   7.0

Searched:        2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters:      4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09902432/runat_12122002_141451_25800/app_query.fasta_1.1735
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09902432_@CGN_1_1_6311_@runat_12122002_141451_25800 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :       GenEmbl:*
                 1:  gb_ba:*
                 2:  gb_htg:*
                 3:  gb_in:*
                 4:  gb_om:*
                 5:  gb_ov:*
                 6:  gb_pat:*
                 7:  gb_ph:*
                 8:  gb_pl:*
                 9:  gb_pr:*
                 10:  gb_ro:*
                 11:  gb_sts:*
                 12:  gb_sy:*
                 13:  gb_un:*
                 14:  gb_vi:*
                 15:  em_ba:*
                 16:  em_fun:*
                 17:  em_hum:*
                 18:  em_in:*
                 19:  em_mu:*
                 20:  em_om:*
                 21:  em_or:*
                 22:  em_ov:*
                 23:  em_pat:*
                 24:  em_ph:*
                 25:  em_pl:*
                 26:  em_ro:*
                 27:  em_sts:*
                 28:  em_un:*
                 29:  em_vi:*
                 30:  em_htg_hum:*
                 31:  em_htg_inv:*
                 32:  em_htg_other:*
                 33:  em_htg_mus:*
                 34:  em_htg_pln:*
                 35:  em_htg_rod:*
                 36:  em_htg_mam:*
                 37:  em_htg_vrt:*
                 38:  em_sy:*
                 39:  em_htgo_hum:*
                 40:  em_htgo_mus:*
                 41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8073	100.0	6160	10	RNU23146	U23146 Rattus norv
	2	8030	99.5	5236	10	RNU41453	U41453 Rattus norv
	3	7461	92.4	178954	2	AC115647	AC115647 Rattus no
	4	6663	82.5	6195	10	AB020886	AB020886 Mus muscu
	5	6663	82.5	6197	10	AF326228	AF326228 Mus muscu
	6	6255	77.5	5902	10	AF326230	AF326230 Mus muscu
	7	6005	74.4	5134	6	AR071282	AR071282 Sequence
	8	4340.5	53.8	6614	6	AX405857	AX405857 Sequence
	9	4300.5	53.3	6608	6	AX329766	AX329766 Sequence
	10	4300.5	53.3	6608	6	AX410793	AX410793 Sequence
	11	4300.5	53.3	6608	9	HSU81607	U81607 Homo sapien
	12	4281	53.0	6605	6	AR002603	AR002603 Sequence
	13	4231.5	52.4	6297	9	AF001504	AF001504 Homo sapi
	14	4097	50.7	104939	9	AL590413	AL590413 Human DNA
	15	4097	50.7	171158	2	AC023201	AC023201 Homo sapi
	16	4090	50.7	6287	9	AB003476	AB003476 Homo sapi
	17	1805	22.4	8439	5	AF308810	AF308810 Xenopus l
c	18	1585.5	19.6	102822	2	AC097731	AC097731 Rattus no
	19	1067	13.2	1027	9	BC000188	BC000188 Homo sapi
	20	1067	13.2	1027	9	BC022814	BC022814 Homo sapi
	21	642	8.0	160827	3	AC099025	AC099025 Drosophil
	22	639.5	7.9	162921	3	AC007853	AC007853 Drosophil
	23	639.5	7.9	181132	3	AC008206	AC008206 Drosophil
	24	639.5	7.9	227398	3	AE003750	AE003750 Drosophil
	25	635	7.9	16377	3	AF202180	AF202180 Plasmodiu
c	26	634	7.9	193262	3	AC007579	AC007579 Drosophil
c	27	634	7.9	260367	3	AE003808	AE003808 Drosophil
	28	629	7.8	113474	2	AC013932	AC013932 Drosophil
	29	612.5	7.6	134855	2	AP004077	AP004077 Oryza sat
	30	601.5	7.5	32009	3	CET06D8	Z49130 Caenorhabdi
	31	589.5	7.3	53424	3	AB055861	AB055861 Procambar
	32	583.5	7.2	75313	2	AC017682	AC017682 Drosophil
c	33	582	7.2	144056	3	AC091222	AC091222 Drosophil
	34	582	7.2	194006	3	AC010063	AC010063 Drosophil
	35	577.5	7.2	39749	3	DMC49E4	AL031128 Drosophil
c	36	576.5	7.1	179145	9	AC009785	AC009785 Homo sapi
	37	576.5	7.1	181360	3	AC104141	AC104141 Drosophil
	38	565.5	7.0	302855	3	AE003420	AE003420 Drosophil
	39	562.5	7.0	68727	3	AC004516	AC004516 Drosophil
	40	557	6.9	71023	2	AC004426	AC004426 Drosophil
c	41	547.5	6.8	23554	3	AC024760	AC024760 Caenorhab
c	42	537	6.7	304106	3	AE003473	AE003473 Drosophil
	43	536	6.6	70398	3	DME271740	AJ271740 Drosophil
c	44	533	6.6	125398	2	AP004150	AP004150 Oryza sat
	45	517	6.4	73391	8	AP002057	AP002057 Arabidops

ALIGNMENTS

RESULT 1

```
RNU23146
LOCUS       RNU23146              6160 bp    DNA     linear   ROD 19-OCT-1995
DEFINITION  Rattus norvegicus mitogenic regulation SSeCKS (322) gene, complete
            cds.
ACCESSION   U23146
VERSION     U23146.1  GI:1022682
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 6160)
  AUTHORS   Lin,X., Nelson,P.J., Frankfort,B., Tombler,E., Johnson,R. and
            Gelman,I.H.
  TITLE     Isolation and characterization of a novel mitogenic regulatory
            gene, 322, which is transcriptionally suppressed in cells
            transformed by src and ras
  JOURNAL   Mol. Cell. Biol. 15 (5), 2754-2762 (1995)
  MEDLINE   95257957
   PUBMED   7739556
REFERENCE   2  (bases 1 to 6160)
  AUTHORS   Gelman,I.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-MAR-1995) Irwin H. Gelman, Microbiology, Mount Sinai
            School of Medicine, One Gustave L. Levy Place, New York, NY 10029,
            USA
COMMENT     On Oct 19, 1995 this sequence version replaced gi:733449.
FEATURES             Location/Qualifiers
     source          1. .6160
                     /organism="Rattus norvegicus"
                     /strain="Fisher"
                     /db_xref="taxon:10116"
                     /clone="13.2.2"
                     /cell_line="Rat-6"
                     /cell_type="fibroblast"
                     /tissue_type="embryo"
                     /clone_lib="R1-a (Robert Kraus, Mount Sinai School of
                     Medicine)"
     5'UTR           1. .32
     gene            33. .5044
                     /gene="322"
     CDS             33. .4823
                     /gene="322"
                     /function="possible mitogenic regulatory gene"
                     /note="Src Suppressed C Kinase Substrate"
                     /citation=[1]
                     /codon_start=1
                     /evidence=experimental
                     /product="SSeCKS"
                     /protein_id="AAA79517.1"
                     /db_xref="GI:1022683"
                     /translation="MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPAD
                     ADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR
                     VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKF
                     VGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESE
                     LKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPV
                     NSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPA
                     SEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQE
                     VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELM
                     KSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSS
                     SGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL
                     EKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEK
                     VKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA
                     SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEG
                     EGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRR
                     KKRADGKQEQATVEDSGPVEINEDDPNVPAVVPLSEYNAVEREKMEAQGNTELPQLLG
                     AVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVE
                     EVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEA
                     SGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK
                     EESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGS
                     ETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLA
                     DSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGE
                     EPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKA
                     ADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA
                     APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQT
                     AEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAE
                     SIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEP
                     EILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPACRLDSREPNRCWTK
                     MKDAKMKHPVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK"
     misc_signal     1581. .1590
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     misc_signal     1797. .1811
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     misc_signal     2061. .2073
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     misc_signal     2301. .2313
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     misc_binding    4230. .4358
                     /gene="322"
                     /note="encodes zinc-finger domain"
                     /citation=[1]
                     /bound_moiety="DNA, RNA or protein binding domain"
     3'UTR           4816. .6160
     misc_signal     5032. .5044
                     /gene="322"
                     /note="encodes nuclear localization signal; NLS"
                     /citation=[1]
     polyA_signal    5760. .5765
     polyA_signal    6136. .6141
     polyA_site      6160
BASE COUNT     1853 a   1409 c   1753 g   1145 t
ORIGIN

Alignment Scores:
Pred. No.:              3.18e-280      Length:        6160
Score:                  8073.00        Matches:       1596
Percent Similarity:     100.00%        Conservative:  0
Best Local Similarity:  100.00%        Mismatches:    0
Query Match:            100.00%        Indels:        0
DB:                     10             Gaps:          0

US-09-902-432-4 (1-1596) x RNU23146 (1-6160)

Qy        1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92

Qy       21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy       41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy       61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy       81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy      101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392
```

```
Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy    341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy    361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy    381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy    401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy    421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy    441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy    461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy    481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532

Qy    501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592

Qy    521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652

Qy    541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy    561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy    581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy    601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy    621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy    641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy    661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy    681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132

Qy    701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy    741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy    761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2433 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2493 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2552

Qy    841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   2553 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612

Qy    861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy    881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy    901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy    921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy    941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy    961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy    981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy   1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy   1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy   1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212

Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812

Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy   1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy   1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy   1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy   1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy   1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy   1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy   1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy   1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy   1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4652

Qy   1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4712

Qy   1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772
```

```
Qy   1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820

RESULT 2
RNU41453
LOCUS       RNU41453                5236 bp    mRNA    linear   ROD 12-JAN-1999
DEFINITION  Rattus norvegicus PKC binding protein and substrate mRNA, complete
            cds.
ACCESSION   U41453
VERSION     U41453.1  GI:4096819
KEYWORDS    .
SOURCE      Rattus norvegicus.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 5236)
  AUTHORS   Chapline,C., Mousseau,B., Ramsay,K. and Jaken,S.
  TITLE     Identification of a major protein kinase C binding protein and
            substrate in rat embryo fibroblasts: Decreased expression in
            transformed cells
  JOURNAL   J. Biol. Chem. (1995) In press
REFERENCE   2  (bases 1 to 5236)
  AUTHORS   Chapline,C., Mousseau,B., Ramsay,K. and Jaken,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-NOV-1995) Susan Jaken, W Alton Jones Cell Science
            Center, Old Barn Road, Lake Placid, NY 12946, USA
FEATURES             Location/Qualifiers
     source          1..5236
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
                     /clone="72"
     CDS             33..4823
                     /note="major PKC binding protein in REF52 cells"
                     /codon_start=1
                     /product="PKC binding protein and substrate"
                     /protein_id="AAD03788.1"
                     /db_xref="GI:4096820"
                     /translation="MGAGSSTEQRSPEQPAGSDTPSELVLSGHGPAAEASGAAGDPAD
                     ADPATKLPQKNGQLSSVNGVAEQGDVHVQEENQEGQEEEVVDEDVGQRESEDVREKDR
                     VEEMAANSTAVEDITKDGQEETSEIIEQIPASENNVEEMVQPAESQANDVGFKKVFKF
                     VGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPSVETAVGESASKESE
                     LKQSTEKQEGTLKQEQSSTEIPLQAESDQAAEEEAKDEGEEKQEKEPTKSPESPSSPV
                     NSETTSSFKKFFTHGWAGWRKKTSFKKSKEDDLETAEKRKEQEAEKVDEEEKEKTEPA
                     SEEQEPAEDTDQARLSADYEKVELPLEDQVGDLEASSEEKCAPLATEVFDEKMEAHQE
                     VVAEVHVSTVEKTEEEQGGGGEAEGGVVVEGTGESLPPEKLAEPQEVPQEAEPAEELM
                     KSREMCVSGGDHTQLTDLSPEEKTLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSS
                     SGLKKLSGKKQKGKRGGGGDEEPGEYQHIHTESPESADEQKGESSASSPEEPEETTCL
                     EKGPLEAPQDGEAEEGTTSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELEK
                     VKSATLSSTDSTVSEMQDEVKTVGEEQKPEEPKRRVDTSVSWEALICVGSSKKRARKA
                     SSSDDEGGPRTLGGDSHRAEEASKDKEAGTDAVPASTQEQDQAQGSSSPEPAGSPSEG
                     EGVSTWESFKRLVTPRKKSKSKLEEKAEDSSVEQLSTEIEPSREESWVSIKKFIPGRR
                     KKRADGKQEQATVEDSGPVEINEDDPDVPAVVPLSEYDAVEREKMEAQGNAELPQLLG
                     AVYVSEELSKTLVHTVSVAVIDGTRAVTSVEERSPSWISASVTEPLEHTAGEAMPPVE
                     EVTEKDIIAEETPVLTQTLPEGKDAHDDMVTSEVDFTSEAVTATETSEALRTEEVTEA
                     SGAEETTDMVSAVSQLTDSPDTTEEATPVQEVESGVLDTEEEERQTQAILQAVADKVK
                     EESQVPATQTVQRTGSKALEKVEEVEEDSEVLASEKEKDVMPKGPVQEAGAEHLAQGS
                     ETGQATPESLEVPEVTADVDHVATCQVIKLQQLMEQAVAPESSETLTDSETNGSTPLA
                     DSDTADGTQQDETIDSQDSKATAAVRQSQVTEEEAATAQKEEPSTLPNNVPAQEEHGE
                     EPGRDVLEPTQQELTAAAVPVLAKTEVGQEGEVDWLDGEKVKEEQEVFVHSGPNSQKA
                     ADVTYDSEVMGVAGCQEKESTEVQSLSLEEGEMETDVEKEKRETKPEQVSEEGEQETA
                     APEHEGTYGKPVLTLDMPSSERGKALGSLGGSPSLPDQDKAGCIEVQVQSLDTTVTQT
                     AEAVEKVIETVVISETGESPECVGAHLLPAEKSSATGGHWTLQHAEDTVPLGPESQAE
                     SIPIIVTPAPESTLHPDLQGEISASQRERSEEEDKPDAGPDADGKESTAIEKVLKAEP
                     EILELESKSNKIVLNVIQTAVDQFARTETAPETHAYDSQTQVPAMQADSREPNRCWTK
                     MKVAKMKHPVPQPREDLQVLTVLEAWAQPRKCLPRLQLKAPVSK"
BASE COUNT     1622 a   1206 c   1583 g    825 t
ORIGIN

Alignment Scores:
Pred. No.:              9.42e-279      Length:        5236
Score:                  8030.00        Matches:       1589
Percent Similarity:          99.75%        Conservative:    3
Best Local Similarity:       99.56%        Mismatches:      4
Query Match:                 99.47%        Indels:          0
DB:                          10            Gaps:            0

US-09-902-432-4 (1-1596) x RNU41453 (1-5236)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92

Qy     21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy     41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy     61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy     81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy    101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db     993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy     341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy     361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy     381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy     401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy     421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy     441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy     461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy     481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532

Qy     501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1592

Qy     521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652

Qy     541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy     561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy     581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy     601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy     621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy     641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy     661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy     681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132

Qy     701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy     721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy     741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy     761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy     781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy     801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2433 AGGGCAGATGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy     821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
           ||||||||||:::||||||||||||||||||||||||||||||:::||||||||||||||
Db    2493 GAGGACGACCCTGATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGG 2552

Qy     841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
           ||||||||||||||||||||||     |||||||||||||||||||||||||||||||||
Db    2553 GAGAAGATGGAAGCCCAGGGGAATGCGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612

Qy     861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy     881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy     901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy     921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy     941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy     961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy     981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy    1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy    1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy    1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212
```

```
Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812

Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy   1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy   1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy   1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy   1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy   1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy   1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy   1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy   1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy   1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||  :::
Db   4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCAG 4652

Qy   1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
             ||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db   4653 GCTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACAC 4712

Qy   1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772

Qy   1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820
```

```
RESULT 3
AC115647
LOCUS       AC115647               178954 bp    DNA     linear   HTG 13-JUL-2002
DEFINITION  Rattus norvegicus clone CH230-171G16, *** SEQUENCING IN PROGRESS
            ***, 53 unordered pieces.
ACCESSION   AC115647
VERSION     AC115647.3  GI:21737529
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 178954)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
```

```
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178954)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (22-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 178954)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jul 12, 2002 this sequence version replaced gi:21239989.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: GPHP
                Center clone name: CH230-171G16
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 142023 bases at least Q40
                Consensus quality: 146696 bases at least Q30
                Consensus quality: 150215 bases at least Q20
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 53 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1459: contig of 1459 bp in length
            *     1460      1559: gap of unknown length
            *     1560      2931: contig of 1372 bp in length
            *     2932      3031: gap of unknown length
            *     3032      4420: contig of 1389 bp in length
            *     4421      4520: gap of unknown length
            *     4521      5800: contig of 1280 bp in length
            *     5801      5900: gap of unknown length
            *     5901      7559: contig of 1659 bp in length
            *     7560      7659: gap of unknown length
            *     7660      8887: contig of 1228 bp in length
            *     8888      8987: gap of unknown length
            *     8988     10297: contig of 1310 bp in length
            *    10298     10397: gap of unknown length
            *    10398     12090: contig of 1693 bp in length
            *    12091     12190: gap of unknown length
            *    12191     13522: contig of 1332 bp in length
            *    13523     13622: gap of unknown length
            *    13623     15629: contig of 2007 bp in length
            *    15630     15729: gap of unknown length
            *    15730     17540: contig of 1811 bp in length
            *    17541     17640: gap of unknown length
            *    17641     19819: contig of 2179 bp in length
            *    19820     19919: gap of unknown length
            *    19920     21261: contig of 1342 bp in length
            *    21262     21361: gap of unknown length
            *    21362     23785: contig of 2424 bp in length
            *    23786     23885: gap of unknown length
            *    23886     25562: contig of 1677 bp in length
            *    25563     25662: gap of unknown length
            *    25663     27442: contig of 1780 bp in length
            *    27443     27542: gap of unknown length
            *    27543     30415: contig of 2873 bp in length
            *    30416     30515: gap of unknown length
            *    30516     31784: contig of 1269 bp in length
            *    31785     31884: gap of unknown length
            *    31885     34970: contig of 3086 bp in length
            *    34971     35070: gap of unknown length
            *    35071     36233: contig of 1163 bp in length
            *    36234     36333: gap of unknown length
            *    36334     39668: contig of 3335 bp in length
            *    39669     39768: gap of unknown length
            *    39769     42866: contig of 3098 bp in length
            *    42867     42966: gap of unknown length
            *    42967     46088: contig of 3122 bp in length
            *    46089     46188: gap of unknown length
            *    46189     49027: contig of 2839 bp in length
            *    49028     49127: gap of unknown length
            *    49128     51579: contig of 2452 bp in length
            *    51580     51679: gap of unknown length
            *    51680     55192: contig of 3513 bp in length
            *    55193     55292: gap of unknown length
            *    55293     58204: contig of 2912 bp in length
            *    58205     58304: gap of unknown length
            *    58305     60939: contig of 2635 bp in length
            *    60940     61039: gap of unknown length
            *    61040     64184: contig of 3145 bp in length
            *    64185     64284: gap of unknown length
            *    64285     67776: contig of 3492 bp in length
            *    67777     67876: gap of unknown length
            *    67877     72019: contig of 4143 bp in length
            *    72020     72119: gap of unknown length
            *    72120     75147: contig of 3028 bp in length
            *    75148     75247: gap of unknown length
            *    75248     78405: contig of 3158 bp in length
            *    78406     78505: gap of unknown length
            *    78506     82238: contig of 3733 bp in length
            *    82239     82338: gap of unknown length
            *    82339     85153: contig of 2815 bp in length
            *    85154     85253: gap of unknown length
            *    85254     87767: contig of 2514 bp in length
            *    87768     87867: gap of unknown length
            *    87868     90643: contig of 2776 bp in length
            *    90644     90743: gap of unknown length
            *    90744     94013: contig of 3270 bp in length
            *    94014     94113: gap of unknown length
            *    94114     98293: contig of 4180 bp in length
            *    98294     98393: gap of unknown length
            *    98394    101682: contig of 3289 bp in length
            *   101683    101782: gap of unknown length
            *   101783    104237: contig of 2455 bp in length
            *   104238    104337: gap of unknown length
            *   104338    108706: contig of 4369 bp in length
            *   108707    108806: gap of unknown length
            *   108807    113208: contig of 4402 bp in length
            *   113209    113308: gap of unknown length
            *   113309    117238: contig of 3930 bp in length
            *   117239    117338: gap of unknown length
            *   117339    122577: contig of 5239 bp in length
```

```
          *    122578    122677: gap of unknown length
          *    122678    126779: contig of 4102 bp in length
          *    126780    126879: gap of unknown length
          *    126880    131807: contig of 4928 bp in length
          *    131808    131907: gap of unknown length
          *    131908    138315: contig of 6408 bp in length
          *    138316    138415: gap of unknown length
          *    138416    143597: contig of 5182 bp in length
          *    143598    143697: gap of unknown length
          *    143698    152755: contig of 9058 bp in length
          *    152756    152855: gap of unknown length
          *    152856    157897: contig of 5042 bp in length
          *    157898    157997: gap of unknown length
          *    157998    167260: contig of 9263 bp in length
          *    167261    167360: gap of unknown length
          *    167361    178954: contig of 11594 bp in length.
FEATURES             Location/Qualifiers
     source          1..178954

Alignment Scores:
Pred. No.:              4.88e-257     Length:         178954
Score:                  7461.00       Matches:        1488
Percent Similarity:     99.20%        Conservative:   6
Best Local Similarity:  98.80%        Mismatches:     11
Query Match:            92.42%        Indels:         3
DB:                     2             Gaps:           0

US-09-902-432-4 (1-1596) x AC115647 (1-178954)

Qy      89 ValGlyGlnArgGluSerGluAspValArgGluLysAspArgValGluGluMetAlaAla 108
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174384 GTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAAGACCGAGTTGAAGAAATGGCGGCC 174443

Qy     109 AsnSerThrAlaValGluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIle 128
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174444 AACTCCACAGCTGTTGAAGATATCACAAAGGATGGCAGGAGGAGACATCAGAAATAATT 174503

Qy     129 GluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnProAlaGluSerGln 148
           |||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db  174504 GAACAGATCCCTGCTTCAGAAAGCAATGTGGAAGAAATGGTACAGCCTGCTGAGTCCCAG 174563

Qy     149 AlaAsnAspValGlyPheLysLysValPheLysPheValGlyPheLysPheThrValLys 168
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174564 GCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAG 174623

Qy     169 LysAspLysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGly 188
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174624 AAGGATAAAAATGAAAAGTCAGATACTGTCCAACTACTCACTGTCAAGAAGGATGAAGGC 174683

Qy     189 GluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAla 208
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174684 GAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCACCAGGAGCCCAGTGTGGAGACTGCC 174743

Qy     209 ValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGly 228
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174744 GTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGC 174803

Qy     229 ThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAla 248
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174804 ACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCCCTTCAAGCCGAATCTGATCAAGCG 174863

Qy     249 AlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSer 268
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174864 GCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAACAAGAGAAAGAGCCCACCAAGTCC 174923

Qy     269 ProGluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePhe 288
           ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db  174924 CCAGAATCCCCGAGCAGCCCAGTCAGCAGTGAGACAACATCTTCCTTCAAGAAGTTCTTC 174983

Qy     289 ThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAsp 308
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  174984 ACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAATCAAAAGAGGATGAT 175043

Qy     309 LeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLys 328
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175044 CTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCAGAAAAAGTAGACGAGGAAGAAAAG 175103

Qy     329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175104 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 175163

Qy     349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175164 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 175223

Qy     369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175224 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 175283

Qy     389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175284 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 175343

Qy     409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175344 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 175403

Qy     429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175404 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 175463

Qy     449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175464 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 175523

Qy     469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175524 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 175583

Qy     489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175584 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 175643

Qy     509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175644 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 175703

Qy     529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175704 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 175763

Qy     549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175764 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 175823

Qy     569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175824 GGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAGGAAGGAACTACTTCCGATGGAGAG 175883

Qy     589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175884 AAGAAGAGGGAAGGGATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 175943

Qy     609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  175944 CGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 176003

Qy     629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  176004 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 176063

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  176064 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 176123

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688
```

```
Db 176124 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATGATGAA 176183
Qy     689 GlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLys 708
Db 176184 GGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAAA 176243
Qy     709 GluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySer 728
Db 176244 GAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAGT 176303
Qy     729 SerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSer 748
Db 176304 TCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTCA 176363
Qy     749 PheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGlu 768
Db 176364 TTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGCCGAA 176423
Qy     769 AspSerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db 176424 GACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAACCGAGTAGAGAAGAATCTTGGGTT 176483
Qy     789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db 176484 TCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGAAGCAAGAACAA 176543
Qy     809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db 176544 GCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGACCCTGATGTCCCAGCC 176603
Qy     829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsn 848
Db 176604 GTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAGCCCAGGGGAAT 176663
Qy     849 ThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeu 868
Db 176664 GCGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTGTCCGAGGAGCTCAGTAAGACTCTG 176723
Qy     869 ValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGlu 888
Db 176724 GTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCAGTGTCGAAGAG 176783
Qy     889 ArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGlu 908
Db 176784 CGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACACAGCGGGAGAA 176843
Qy     909 AlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVal 928
Db 176844 GCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAGAAACTCCTGTG 176903
Qy     929 LeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluVal 948
Db 176904 CTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCACCAGTGAAGTG 176963
Qy     949 AspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGlu 968
Db 176964 GATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCCGTACTGAAGAA 177023
Qy     969 ValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGlnLeu 988
Db 177024 GTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAGTTTCCCAGCTG 177083
Qy     989 ThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGluSerGlyVal 1008
Db 177084 ACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGGGTGGTGTG 177143
Qy    1009 LeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLys 1028
Db 177144 CTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCGTTGCAGACAAG 177203
Qy    1029 ValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrGlySerLysAla 1048
Db 177204 GTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGGGGTCAAAAGCA 177263
Qy    1049 LeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGluLysGluLys 1068
Db 177264 CTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGGAGAAAGAGAAG 177323
Qy    1069 AspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySer 1088
Db 177324 GACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTGCACAGGGCTCT 177383
Qy    1089 GluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrAlaAspValAsp 1108
Db 177384 GAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAGCAGATGTAGAC 177443
Qy    1109 HisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnAlaValAlaPro 1128
Db 177444 CATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGGCCGTGGCCCCT 177503
Qy    1129 GluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAspSer 1148
Db 177504 GAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCTTAGCAGATTCA 177563
Qy    1149 AspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAlaThr 1168
Db 177564 GACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACAGTAAAGCCACT 177623
Qy    1169 AlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGlu 1188
Db 177624 GCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTCAGAAAGAGGAG 177683
Qy    1189 ProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGlyArg 1208
Db 177684 CCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAGAACCAGGAAGA 177743
Qy    1209 AspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLys 1228
Db 177744 GATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCGTTCTGGCAAAG 177803
Qy    1229 ThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGluGlu 1248
Db 177804 ACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAGTCAAAGAAGAA 177863
Qy    1249 GlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAsp 1268
Db 177864 CAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATGTGACATATGAC 177923
Qy    1269 SerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSerLeu 1288
Db 177924 AGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAGTGCAGAGTCTT 177983
Qy    1289 SerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysPro 1308
Db 177984 AGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGGAGACAAAGCCA 178043
Qy    1309 GluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyr 1328
Db 178044 GAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATGAAGGAACCTAC 178103
Qy    1329 GlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySer 1348
Db 178104 GGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGGCACTGGGAAGC 178163
Qy    1349 LeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnVal 1368
Db 178164 CTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAGAGGTTCAAGTT 178223
Qy    1369 GlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThr 1388
Db 178224 CAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGGTCATAGAAACG 178283
Qy    1389 ValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAla 1408
Db 178284 GTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACTTATTACCAGCT 178343
```

```
Qy    1409 GluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValPro 1428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 178344 GAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGGACACGGTACCC 178403

Qy    1429 LeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSer 1448
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 178404 CTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTGCTCCTGAAAGC 178463

Qy    1449 ThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGlu 1468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 178464 ACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGCGATCAGAGGAA 178523

Qy    1469 GluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLys 1488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 178524 GAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAGCAATCGAAAAA 178583

Qy    1489 ValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 178584 GTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACAAGATTGTGCTG 178643

Qy    1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaProGluThrHis 1528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 178644 AACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCCCCGAAACTCAT 178703

Qy    1529 AlaTyrAspSerGlnThrGlnValProAla-CysArgLeuAspSerArgGluProAsnAr 1548
           |||||||||||||||||||||||||||||| ||||||||    :::||||||||||||||
Db 178704 GCTTATGATTCACAGACCCAGGTTCCTGCAATGCAGGCTGACAGCCAGGGAGCCCAACAG 178763

Qy    1548 gCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnProArgGluAs 1568
           ||||||||||||||||||    ||||||||||||||||||||||||||||||||||||||
Db 178764 ATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGCCCAGAGAGGA 178823

Qy    1568 pLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLeuProArgLe 1588
           ||||||||||||||||||||||||   ||||||||||||   :::||||||       ||
Db 178824 CTTGCAAGTCCTGACCGTTCTGGAGCA-TGGGCTCAGCCTCC-GAATGCTTGGCCGCGCT 178881

Qy    1588 uGlnLeuLysAlaPro 1593
           |      ||||||||
Db 178882 GCANNTGAAAGCGCCG 178897
```

```
RESULT 4
AB020886
LOCUS       AB020886                6195 bp    mRNA    linear   ROD 14-APR-1999
DEFINITION  Mus musculus mRNA for SSeCKS, complete cds.
ACCESSION   AB020886
VERSION     AB020886.1  GI:4589831
KEYWORDS    SSeCKS.
SOURCE      Mus musculus (strain:BALB/c, isolate:200) 9-12 weeks male testis
            tissue_lib:Clontech mouse testis library cDNA to mRNA, clone:133.5.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (sites)
  AUTHORS   Okita,K., Kitamura,H., Iwanaga,T., Morimatsu,M. and Saito,M.
  TITLE     Isolation and characterization of mouse SSeCKS
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6195)
  AUTHORS   Okita,K., Kitamura,H., Saito,M. and Morimatsu,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1998) Keisuke Okita, School of Veterinary
            Medicine, Hokkaido University, Laboratory of Biochemistry; N18, W9,
            Kita-ku, Sapporo, Hokkaido 060-0818, Japan
            (E-mail:v080008@vet.ec.hokudai.ac.jp, Tel:81-11-706-5206,
            Fax:81-11-757-0703)
COMMENT     Sequence updated (10-Mar-1999).
FEATURES             Location/Qualifiers
     source          1..6195
                     /organism="Mus musculus"
                     /strain="BALB/c"
                     /isolate="200"
                     /db_xref="taxon:10090"
                     /clone="133.5"
                     /sex="male"
                     /tissue_type="testis"
                     /dev_stage="9-12 weeks"
                     /tissue_lib="Clontech mouse testis library"
     gene            1..6195
                     /gene="ssecks"
     CDS             140..5194
                     /gene="ssecks"
                     /note="Src Suppressed C Kinase Substrate"
                     /codon_start=1
                     /product="SSeCKS"
                     /protein_id="BAA76894.1"
                     /db_xref="GI:4589832"
                     /translation="MGAGSSTEQRSPEQPAESDTPSELELSGHGPAAEASGAAGDPAD
                     ADPATKLPQKNGQLSAVNGVAEQEDVHVQEESQDGQEEEVTVEDVGQRESEDVKEKDR
                     AKEMAASSTVVEDITKDEQEETPEIIEQIPASESNVEEMAQAAESQANDVGFKKVFKF
                     VGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPGVETVGESASKESEL
                     KQSTEKQEGTLKQAQSSTEIPLQAESGQGTEEEAAKDGEENREKEPTKPLESPTSPVS
                     NETTSSFKKFFTHGWAGWRKKTSFKKPKEDDLETSEKRKEQEAEKVDEEEGEKTEPAP
                     AEEQEPAEGTDQARLSADYEKVELPLEDQVGDLEALSEKCAPLATEVFDEKTEAHQEV
                     VAEVHVSTVEKMTKGQGGAEVEGDVVVEGSGESLPPEKLAETQEVPQEAEPVEELMKT
                     KEVCVSGGDHTQLTDLSPEEKMLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSG
                     LKKLSGKKQKGKRGGGGGDEEPGEYQHIQTESPESADEQKGESSASSPEEPEEIACLE
                     KGPSEAPQEAEAEEGATSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELDKV
                     KSATLSSTESTASGMQDEVRAVGEEQRSEEPKRRVDTSVSWEALICVGSSKKRARKAS
                     SSDDEGGPRTLGGDGHRAEEASKDKEADALPASTQEQDQAHGSSSPEPAGSPSEGEGV
                     STWESFKRLVTPRKKSKSKLEERAEDSGAEQLASEIEPSREESWVSIKKFIPGRRKKR
                     ADGKQEQAAVEDSGPGEINEDDPDVPAVVPLSEYDAVEREKLEAQRAQENVELPQLKG
                     AVYVSEELSKTLVHTVSVAVIDGTRAVTSAEERSPSWISASMTEPLEHAEGVATPPVG
                     EVTEKDITAEATPALAQTLPGGKDAHDDIVTSEVDFTSEAVTAAETTEALRAEELTEA
                     SGAEETTDMVSAVSQLSDSPDTTEEATPVQEVEGGMLDTEEQERQTQAVLQAVADKVK
                     EDSQVPATQTLQRAGPKALEKVEEVEEDSEVLATEKEKDVVPEGPVQEAETEHLAQGS
                     ETVQATPESLEVPEVTEDVDRATTCQVIKHQQLMEQAVAPESSETLTDSETNGSTPLA
                     DSDTPNGTQQDETVDSQDSNAIAAVKQSQVTEEEAAAAQTEGPSTPSSFPAQEEHREK
                     PGRDVLEPTQALAAGAVPILAKAEVGQEGEAGQFDGEKVKDGQCVKELEVPVHTGPNS
                     QKTADLTRDSEVMEVARCQETESNEEQSISPEKREMGTDVEKEETETKTEQASEEHEQ
                     ETAAPEHEGTHPKPVLTADMPHSERGKALGSLEGSPSLPDQDKADCIEVQVQSSDTPV
                     TQTTEAVKKVEETVATSEMDESLECAGAQSLPAEKLSETGGYGTLQHGEDTVPQGPES
                     QAESIPIIVTPAPESILHSDLQREVSASQKQRSDEDNKPDAGPDAAGKESAAREKILR
                     AEPEILELESKSNKIVQSVIQTAVDQFARTETAPETHASDLQNQVPVMQADSQGAQQM
                     LDKDESDLQVSPQDGTLSAVAQEGLAVSDSSEGMSKASEMITTLAVESASVKESVEKL
                     PLQCKDEKEHAADGPQHQSLAKAEADASGNLTKESPDTNGPKLTEEGDALKEEMNKAQ
                     TEEDDLQEPKGDLTES"
     polyA_site      6195
                     /gene="ssecks"
                     /note="20 A nucleotides"
BASE COUNT     1838 a   1451 c   1777 g   1129 t
ORIGIN

Alignment Scores:
Pred. No.:              7.31e-230     Length:          6195
Score:                  6663.00       Matches:         1357
Percent Similarity:     88.90%        Conservative:    76
Best Local Similarity:  84.18%        Mismatches:      156
Query Match:            82.53%        Indels:          23
DB:                     10            Gaps:            11

US-09-902-432-4 (1-1596) x AB020886 (1-6195)

Qy       1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
           ||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db     140 ATGGGTGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGAGAGCGACACG 199

Qy      21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
           ||||||||||||    ||||||||||||||||||||||||||||||||||||||||||||
Db     200 CCGAGCGAGCTGGAGCTCAGTGGCCATGGGCCCGCAGCGGAAGCGTCGGGAGCAGCTGGA 259

Qy      41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db     260 GATCCCGCTGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGTCAGCTGTCTGCC 319

Qy      61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
           ||||||||||||||||||||   ||||||||||||||||||||:::|||:::||||||
```

```
Db    320 GTCAATGGTGTAGCTGAACAAGAAGATGTCCACGTCCAAGAGGAAAGCCAGGATGGGCAA 379

Qy     81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
Db    380 GAGGAAGAAGTCACTGTTGAAGATGTTGGACAGAGAGAGTCAGAAGATGTGAAAGAAAAA 439

Qy    101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
Db    440 GACCGAGCTAAAGAAATGGCAGCCAGTTCCACAGTTGTTGAAGATATCACAAAGGACGAG 499

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
Db    500 CAGGAGGAAACACCGGAAATAATCGAACAGATCCCTGCTTCAGAGAGCAATGTGGAAGAA 559

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
Db    560 ATGGCGCAGGCTGCTGAGTCCCAAGCTAATGACGTCGGCTTCAAGAAGGTATTTAAATTT 619

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
Db    620 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAACGAAAAGTCAGATACCGTCCAGCTA 679

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
Db    680 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCCGTCGGAGCAGGAGACCAC 739

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
Db    740 CAAGAGCCCGGAGTGGAGACC---GTCGGCGAATCAGCATCCAAAGAAAGTGAGCTGAAG 796

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
Db    797 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGCACAGAGCAGCACAGAAATTCCC 856

Qy    241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
Db    857 CTTCAAGCCGAATCTGGTCAAGGGACCGAGGAAGAAGCAGCCAAAGATGGAGAAGAAAAC 916

Qy    261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
Db    917 CGAGAGAAAGAACCTACCAAGCCCCTAGAATCTCCGACCAGCCCTGTCAGCAATGAGACA 976

Qy    281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
Db    977 ACATCTTCCTTCAAGAAATTCTTCACTCACGGCTGGGCCGGCTGGCGCAAGAAGACCAGC 1036

Qy    301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
Db   1037 TTCAAGAAACCAAAGGAAGATGATCTGGAAACTTCCGAGAAGAGAAAGGAGCAAGAGGCT 1096

Qy    321 GluLysValAspGluGluGluLysGluLysThrGluProAla---SerGluGluGlnGlu 339
Db   1097 GAAAAAGTAGACGAGGAAGAAGGGGAAAAGACAGAGCCAGCCCCAGCCGAGGAGCAGGAG 1156

Qy    340 ProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuPro 359
Db   1157 CCTGCAGAAGGCACAGACCAGGCCAGGTTGTCAGCCGACTATGAGAAGGTGGAGTTGCCT 1216

Qy    360 LeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAla 379
Db   1217 TTGGAAGACCAGGTCGGTGACCTGGAGGCATTGTCG---GAGAAGTGTGCTCCTTTGGCA 1273

Qy    380 ThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisVal 399
Db   1274 ACGGAAGTGTTTGATGAGAAGACGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTG 1333

Qy    400 SerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyVal 419
Db   1334 AGCACCGTGGAGAAGATGACGAAAGGGCAA---GGAGGAGCAGAGGTGGAAGGGGATGTG 1390

Qy    420 ValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluVal 439
Db   1391 GTGGTGGAAGGATCGGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGACCCAGGAGGTC 1450

Qy    440 ProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGly 459
Db   1451 CCCCAGGAAGCTGAGCCTGTGGAGGAGCTGATGAAGACCAAAGAAGTATGCGTCTCTGGG 1510

Qy    460 GlyAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisPro 479
Db   1511 GGTGACCATACTCAGCTGACAGATCTAAGTCCTGAAGAGAAGATGCTACCCAAACACCCC 1570

Qy    480 GluGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGly 499
Db   1571 GAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAGAGAATCAAGGTACAGGGA 1630

Qy    500 SerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGln 519
Db   1631 AGTCCCCTGAAGAAGCTCTTCAGCAGTTCGGGCTTAAAGAAGCTCTCCGGGAAGAAGCAG 1690

Qy    520 LysGlyLysArg---GlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHis 538
Db   1691 AAGGGGAAGAGAGGAGGAGGCGGGGGAGATGAAGAGCCAGGAGAATACCAACACATTCAA 1750

Qy    539 ThrGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGlu 558
Db   1751 ACCGAGTCCCCAGAGAGTGCTGACGAGCAGAAGGGAGAGAGCTCTGCCTCTTCCCCTGAA 1810

Qy    559 GluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGlu 578
Db   1811 GAGCCCGAGGAGATCGCGTGTCTGGAGAAGGGGCCATCGGAAGCACCCCAGGAAGCGGAA 1870

Qy    579 AlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAla 598
Db   1871 GCTGAGGAAGGAGCGACTTCCGACGGAGAGAAGAAAAGGGAAGGGATCACCCCCTGGGCA 1930

Qy    599 SerPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLys 618
Db   1931 TCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGCGACAAG 1990

Qy    619 GluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSer 638
Db   1991 GAAGAAGAGCTGGATAAGGTCAAGAGTGCCACCTTGTCCTCCACGGAGAGCACGGCGTCT 2050

Qy    639 GluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArg 658
Db   2051 GGAATGCAGGATGAGGTCAGAGCGGTTGGCGAGGAGCAAAGGTCAGAGGAGCCAAAGCGC 2110

Qy    659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
Db   2111 AGGGTGGATACTTCAGTGTCTTGGGAGGCGTTGATTTGTGTCGGATCGTCCAAGAAGAGA 2170

Qy    679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
Db   2171 GCGAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGAACACTGGGAGGGGATGGC 2230

Qy    699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
Db   2231 CACAGAGCGGAGGAGGCTAGCAAAGACAAAGAAGCA------GATGCTCTTCCTGCCAGC 2284

Qy    719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
Db   2285 ACCCAGGAACAAGACCAAGCGCACGGAAGTTCCTCACCCGAGCCAGCTGGAAGCCCTTCT 2344

Qy    739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLys 758
Db   2345 GAAGGGGAGGGCGTCTCCACCTGGGAGTCATTTAAGAGATTAGTCACTCCACGAAAAAAA 2404

Qy    759 SerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGlu 778
Db   2405 TCCAAGTCAAAACTGGAAGAGAGAGCCGAAGACTCCGGTGCAGAGCAGTTGGCCTCCGAG 2464

Qy    779 IleGluProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArg 798
Db   2465 ATCGAACCAAGTAGAGAGGAATCTTGGGTTTCCATTAAGAAATTTATTCCTGGACGGCGG 2524
```

```
Qy    799 LysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGlu 818
Db   2525 AAGAAAAGGGCAGATGGAAGCAAGAACAGGCCGCCGTTGAAGACTCGGGGCCAGGAGAG 2584

Qy    819 IleAsnGluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaVal 838
Db   2585 ATCAATGAGGACGACCCCGACGTCCCAGCTGTTGTGCCTCTGTCTGAGTACGATGCGGTA 2644

Qy    839 GluArgGluLysMetGlu---------AlaGlnGlyAsnThrGluLeuProGlnLeuLeu 855
Db   2645 GAGAGAGAGAAGCTGGAAGCGCAGCGAGCTCAGGAGAACGTGGAGCTGCCCCAGCTGAAG 2704

Qy    856 GlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrValSerValAla 875
Db   2705 GGGGCTGTGTATGTGTCTGAGGAGCTTAGTAAGACTCTGGTTCACACCGTGAGTGTCGCG 2764

Qy    876 ValIleAspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSer 895
Db   2765 GTCATTGATGGGACCAGGGCAGTCACCAGTGCCGAAGAGCGGTCCCCTTCGTGGATATCT 2824

Qy    896 AlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGluGlu 915
Db   2825 GCTTCCATGACAGAACCTCTTGAGCACGCAGAGGGAGTGGCCACACCGCCTGTTGGAGAG 2884

Qy    916 ValThrGluLysAspIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGlu 935
Db   2885 GTCACTGAAAAAGACATCACTGCAGAAGCAACTCCTGCACTCGCCCAGACTTTACCAGGG 2944

Qy    936 GlyLysAspAlaHisAspAspMetValThrSerGluValAspPheThrSerGluAlaVal 955
Db   2945 GGCAAAGATGCCCATGACGACATAGTCACCAGTGAAGTGGATTTTACCTCAGAAGCAGTG 3004

Qy    956 ThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAla 975
Db   3005 ACAGCCGCAGAAACCACAGAGGCGCTCCGCGCTGAAGAACTTACCGAAGCATCAGGGGCA 3064

Qy    976 GluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThr 995
Db   3065 GAAGAGACCACAGACATGGTGTCTGCAGTTTCCCAGCTGTCCGACTCCCCGGACACCACA 3124

Qy    996 GluGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGlu 1015
Db   3125 GAGGAAGCCACCCCAGTTCAGGAGGTAGAGGGTGGCATGCTAGATACGGAAGAACAGGAG 3184

Qy   1016 ArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnVal 1035
Db   3185 CGCCAGACGCAGGCCGTCCTCCAAGCCGTTGCAGACAAAGTGAAAGAGGACTCCCAGGTG 3244

Qy   1036 ProAlaThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluVal 1055
Db   3245 CCTGCAACCCAGACTCTGCAGAGAGCAGGGCCGAAAGCACTGGAGAAGGTGGAGGAGGTA 3304

Qy   1056 GluGluAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyPro 1075
Db   3305 GAGGAGGACTCCGAGGTGCTGGCTACCGAGAAAGAGAAGGATGTTGTGCCGGAAGGACCC 3364

Qy   1076 ValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrPro 1095
Db   3365 GTGCAGGAAGCTGAAACTGAGCATCTTGCACAGGGCTCCGAGACTGTACAGGCTACCCCA 3424

Qy   1096 GluSerLeuGluValProGluValThrAlaAspValAspHisValAlaThrCysGlnVal 1115
Db   3425 GAGAGCCTTGAAGTTCCTGAAGTCACAGAGGATGTAGACCGTGCCACCACATGCCAGGTT 3484

Qy   1116 IleLysLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThr 1135
Db   3485 ATCAAGCACCAGCAGCTGATGGAACAGGCTGTGGCCCCTGAGTCATCTGAAACCTTGACA 3544

Qy   1136 AspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGln 1155
Db   3545 GACAGTGAGACAAATGGAAGTACTCCCCTCGCAGATTCAGACACTCCAAACGGGACACAG 3604

Qy   1156 GlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGln 1175
Db   3605 CAAGACGAGACCGTTGACAGCCAGGACAGTAATGCCATTGCCGCCGTCAAGCAGTCACAG 3664

Qy   1176 ValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsn 1195
Db   3665 GTCACTGAAGAGGAGGCAGCTGCTGCTCAGACGGAGGGGCCTTCAACA---CCATCTAGT 3721

Qy   1196 ValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProThrGln 1215
Db   3722 TTTCCAGCCCAGGAAGAACACAGGGAAAAACCAGGAAGGGATGTTCTAGAACCCACA--- 3778

Qy   1216 GlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGly 1235
Db   3779 CAAGCGCTGGCTGCCGGGGCAGTGCCTATTCTGGCAAAGGCTGAGGTGGGTCAAGAGGGT 3838

Qy   1236 GluValAspTrpLeuAspGlyGluLysValLysGlu---------------GluGlnGlu 1250
Db   3839 GAGGCTGGCCAGTTTGATGGAGAAAAAGTCAAAGACGGACAGTGTGTTAAAGAACTGGAG 3898

Qy   1251 ValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGlu 1270
Db   3899 GTGCCTGTGCACACTGGACCCAACAGTCAAAAGACTGCTGACTTGACACGTGACAGTGAA 3958

Qy   1271 ValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSerLeuSerLeu 1290
Db   3959 GTAATGGAAGTGGCCAGATGTCAGGAAACTGAGAGTAATGAAGAACAGAGTATTAGCCCG 4018

Qy   1291 GluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGluGln 1310
Db   4019 GAGAAAAGAGAGATGGGAACCGACGTTGAAAAGGAGGAAACAGAGACCAAGACAGAGCAA 4078

Qy   1311 ValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLys 1330
Db   4079 GCCAGTGAAGAACATGAGCAGGAAACAGCTGCTCCTGAGCATGAAGGAACCCACCCTAAG 4138

Qy   1331 ProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGly 1350
Db   4139 CCAGTCCTGACAGCTGACATGCCTCACTCAGAGAGGGGAAAGGCACTGGGCAGCCTTGAA 4198

Qy   1351 GlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSer 1370
Db   4199 GGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGATTGCATAGAGGTTCAAGTTCAAAGC 4258

Qy   1371 LeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrValVal 1390
Db   4259 TCAGACACACCAGTCACTCAAACAACCGAAGCTGTGAAAAAGGTCGAAGAAACTGTGGCA 4318

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4319 ACTTCAGAGATGGATGAAAGTTTGGAGTGTGCAGGTGCGCAATCATTACCAGCTGAGAAG 4378

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4379 CTCTCCGAAACCGGTGGCTACGGGACTCTTCAGCATGGAGAGGACACCGTGCCCCAGGGG 4438

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4439 CCTGAGTCTCAGGCAGAGTCCATCCCAATAATAGTAACTCCTGCTCCTGAAAGCATCCTA 4498

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4499 CATTCTGACCTTCAAAGAGAAGTGAGCGCATCCCAGAAACAGAGATCAGATGAAGATAAC 4558

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4559 AAGCCAGATGCTGGTCCTGATGCTGCCGGCAAGGAGAGTGCAGCAAGAGAGAAAATCCTC 4618

Qy   1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
Db   4619 AGGGCTGAACCTGAGATCTTGGAACTTGAGAGTAAGAGCAATAAGATTGTCCAGAGTGTC 4678

Qy   1511 IleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaProGluThrHisAlaTyr 1530
```

```
Db     4679 ATCCAGACAGCCGTCGACCAGTTTGCACGTACAGAAACAGCCCCCGAAACCCACGCTTCT 4738

Qy     1531 AspSerGlnThrGlnValProAla-CysArgLeuAspSerArgGluProAsnArgCysTr 1550
            |||   |||   |||||||||   ||||||||||   :::||||||   ||||||||||
Db     4739 GATTTACAGAATCAGGTTCCTGTGATGCAGGCTGACAGCCAGGGAGCACAACAGATGCTG 4798

Qy     1550 pThrLysMetLys---------------AspAlaLysMetLysHisProValProGlnPr 1565
            ||||||||||||                ||||||:::|||  ||||||  ||
Db     4799 GACAAAGATGAAAGCGACCTTCAAGTCTCCCCCAAGATGGAACACTCAGTGCCGTAGCC 4858

Qy     1565 oArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLe 1585
            ||||:::||||||:::   |||  ||||||:::|||  |||:::   :::|||
Db     4859 CAGGAAGGACTTGCGGTTTCTGATAGTTCTGAAGGCATGAGCAAGGCTTCAGAAATGATC 4918

Qy     1585 uProArgLeuGlnLeuLysAlaProValSerLys 1596
             |||||||||||||||||   |||||||||||
Db     4919 ACCACGCTTGCAGTTGAAAGTGCCAGTGTCAAAG 4952

RESULT 5
AF326228
LOCUS       AF326228              6197 bp    mRNA    linear   ROD 16-JUL-2001
DEFINITION  Mus musculus PKC binding protein SSeCKS mRNA, complete cds.
ACCESSION   AF326228
VERSION     AF326228.1  GI:13236403
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 6197)
  AUTHORS   Camus,A., Mesbah,K., Rallu,M., Babinet,C. and Barra,J.
  TITLE     Gene trap insertion reveals two open reading frames in the mouse
            SSeCKS gene: the form predominantly detected in the nervous system
            is suppressed by the insertion while the other, specific of the
            testis, remains expressed
  JOURNAL   Mech. Dev. 105 (1-2), 79-91 (2001)
  MEDLINE   21322683
   PUBMED   11429284
REFERENCE   2  (bases 1 to 6197)
  AUTHORS   Camus,A., Mesbah,K., Babinet,C. and Barra,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2000) Immunologie, Institut Pasteur, 25 rue du
            Docteur Roux, Paris 75015, France
FEATURES             Location/Qualifiers
     source          1..6197
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="10"
     CDS             140..5194
                     /codon_start=1
                     /product="PKC binding protein SSeCKS"
                     /protein_id="AAK16150.1"
                     /db_xref="GI:13236404"
                     /translation="MGAGSSTEQRSPEQPAESDTPSELELSGHGPAAEASGAAGDPAD
                     ADPATKLPQKNGQLSAVNGVAEQEDVHVQEESQDGQEEVTVEDVGQRESEDVKEKDR
                     AKEMAASSTVVEDITKDEQEETPEIIEQIPASESNVEEMAQAAESQANDVGFKKVFKF
                     VGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPGVETVGESASKESEL
                     KQSTEKQEGTLKQAQSSTEIPLQAESGQGTEEAAKDGEENREKEPTKPLESPTSPVS
                     NETTSSFKKFFTHGWAGWRKKTSFKKPKEDDLETSEKRKEQEAEKVDEEEGEKTEPAP
                     AEEQEPAEGTDQARLSADYEKVELPLEDQVGDLEALSEKCAPLATEVFDEKTEAHQEV
                     VAEVHVSTVEKMTKGQGGAEVEGDVVVEGSGESLPPEKLAETQEVPQEAEPVEELMKT
                     KEVCVSGGDHTQLTDLSPEEKMLPKHPEGIVSEVEMLSSQERIKVQGSPLKKLFSSSG
                     LKKLSGKKQKGKRGGGGDEEPGEYQHIQTESPESADEQKGESSASSPEEPEEIACLE
                     KGPSEAPQEAEAEEGATSDGEKKREGITPWASFKKMVTPKKRVRRPSESDKEEELDKV
                     KSATLSSTESTASGMQDEVRAVGEEQRSEEPKRRVDTSVSWEALICVGSSKKRARKAS
                     SSDDEGGPRTLGGDGHRAEEASKDKEADALPASTQEQDQAHGSSSPEPAGSPSEGEGV
                     STWESFKRLVTPRKKSKSKLEERAEDSGAEQLASEIEPSREESWVSIKKFIPGRRKKR
                     ADGKQEQAAVEDSGPGEINEDDPDVPAVVPLSEYDAVEREKLEAQRAQENVELPQLKG
                     AVYVSEELSKTLVHTVSVAVIDGTRAVTSAEERSPSWISASMTEPLEHAEGVATPPVG
                     EVTEKDITAEATPALAQTLPGGKDAHDDIVTSEVDFTSEAVTAAETTEALRAEELTEA
                     SGAEETTDMVSAVSQLSDSPDTTEEATPVQEVEGGMLDTEEQERQTQAVLQAVADKVK
                     EDSQVPATQTLQRAGPKALEKVEEVEEDSEVLATEKEKDVVPEGPVQEAETEHLAQGS
                     ETVQATPESLEVPEVTEDVDRATTCQVIKHQQLMEQAVAPESSETLTDSETNGSTPLA
                     DSDTPNGTQQDETVDSQDSNAIAAVKQSQVTEEEAAAAQTEGPSTPSSFPAQEEHREK
                     PGRDVLEPTQALAAGAVPILAKAEVGQEGEAGQFDGEKVKDGQCVKELEVPVHTGPNS
                     QKTADLTRDSEVMEVARCQETESNEEQSISPEKREMGTDVEKEETETKTEQASEEHEQ
                     ETAAPEHEGTHPKPVLTADMPHSERGKALGSLEGSPSLPDQDKADCIEVQVQSSDTPV
                     TQTTEAVKKVEETVATSEMDESLECAGAQSLPAEKLSETGGYGTLQHGEDTVPQGPES
                     QAESIPIIVTPAPESILHSDLQREVSASQKQRSDEDNKPDAGPDAAGKESAAREKILR
                     AEPEILELESKSNKIVQSVIQTAVDQFARTETAPETHASDLQNQVPVMQADSQGAQQM
                     LDKDESDLQVSPQDGTLSAVAQEGLAVSDSSEGMSKASEMITTLAVESASVKESVEKL
                     PLQCKDEKEHAADGPQHQSLAKAEADASGNLTKESPDTNGPKLTEEGDALKEEMNKAQ
                     TEEDDLQEPKGDLTES"
BASE COUNT     1840 a   1447 c   1779 g   1131 t
ORIGIN

Alignment Scores:
Pred. No.:               7.31e-230      Length:         6197
Score:                   6663.00        Matches:        1357
Percent Similarity:      88.90%         Conservative:   76
Best Local Similarity:   84.18%         Mismatches:     156
Query Match:             82.53%         Indels:         23
DB:                      10             Gaps:           11

US-09-902-432-4 (1-1596) x AF326228 (1-6197)

Qy        1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
            |||||||||||||||||||||||||||||||||||||||||||||||   ||||||||
Db      140 ATGGGTGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGAGAGCGACACG 199

Qy       21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
            ||||||||||||   |||||||||||||||||||||||||||||||||||||||||||||
Db      200 CCGAGCGAGCTGGAGCTCAGTGGCCATGGGCCCGCAGCGGAAGCGTCGGGAGCAGCTGGA 259

Qy       41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db      260 GATCCCGCTGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGTCAGCTGTCTGCC 319

Qy       61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
            ||||||||||||||||||   ||||||||||||||||||:::|||:::||||||
Db      320 GTCAATGGTGTAGCTGAACAAGAAGATGTCCACGTCCAAGAGGAAAGCCAGGATGGGCAA 379

Qy       81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
            ||||||||||||      |||||||||||||||||||||||||||||||||:::||||||
Db      380 GAGGAAGAAGTCACTGTTGAAGATGTTGGACAGAGAGAGTCAGAAGATGTGAAAGAAAAA 439

Qy      101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
            ||||||   :::|||||||||:::||||||||||||||||||||||||
Db      440 GACCGAGCTAAAGAAATGGCAGCCAGTTCCACAGTTGTTGAAGATATCACAAAGGACGAG 499

Qy      121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
            ||||||||||||   ||||||||||||||||||||||||||||||:::||||||||||||
Db      500 CAGGAGGAAACACCGGAAATAATCGAACAGATCCCTGCTTCAGAGAGCAATGTGGAAGAA 559

Qy      141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
            |||   |||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db      560 ATGGCGCAGGCTGCTGAGTCCCAAGCTAATGACGTCGGCTTCAAGAAGGTATTTAAATTT 619

Qy      161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      620 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAACGAAAAGTCAGATACCGTCCAGCTA 679

Qy      181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      680 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCCGTCGGAGCAGGAGACCAC 739

Qy      201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
            |||||||||   |||||||||   |||||||||||||||||||||||||||||||||
Db      740 CAAGAGCCCGGAGTGGAGACC---GTCGGCGAATCAGCATCCAAAGAAAGTGAGCTGAAG 796

Qy      221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
            ||||||||||||||||||||||||||||||||||||   |||||||||||||||||||||
Db      797 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGCACAGAGCAGCACAGAAATTCCC 856

Qy      241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
            |||||||||||||||   |||      ||||||||||||      :::|||||||||
```

```
Db   857 CTTCAAGCCGAATCTGGTCAAGGGACCGAGGAAGAAGCAGCCAAAGATGGAGAAGAAAAC 916

Qy   261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
Db   917 CGAGAGAAAGAACCTACCAAGCCCCTAGAATCTCCGACCAGCCCTGTCAGCAATGAGACA 976

Qy   281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
Db   977 ACATCTTCCTTCAAGAAATTCTTCACTCACGGGTGGGCCGGCTGGCGCAAGAAGACCAGC 1036

Qy   301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
Db  1037 TTCAAGAAACCAAAGGAAGATGATCTGGAAACTTCCGAGAAGAGAAAGGAGCAAGAGGCT 1096

Qy   321 GluLysValAspGluGluGluLysGluLysThrGluProAla---SerGluGluGlnGlu 339
Db  1097 GAAAAAGTAGACGAGGAAGAAGGGGAAAAGACAGAGCCAGCCCCAGCCGAGGAGCAGGAG 1156

Qy   340 ProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuPro 359
Db  1157 CCTGCAGAAGGCACAGACCAGGCCAGGTTGTCAGCCGACTATGAGAAGGTGGAGTTGCCT 1216

Qy   360 LeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAla 379
Db  1217 TTGGAAGACCAGGTCGGTGACCTGGAGGCATTGTCG---GAGAAGTGTGCTCCTTTGGCA 1273

Qy   380 ThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisVal 399
Db  1274 ACGGAAGTGTTTGATGAGAAGACGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTG 1333

Qy   400 SerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyVal 419
Db  1334 AGCACCGTGGAGAAGATGACGAAAGGGCAA---GGAGGAGCAGAGGTGGAAGGGGATGTG 1390

Qy   420 ValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluVal 439
Db  1391 GTGGTGGAAGGATCGGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGACCCAGGAGGTC 1450

Qy   440 ProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGly 459
Db  1451 CCCCAGGAAGCTGAGCCTGTGGAGGAGCTGATGAAGACCAAAGAAGTATGCGTCTCTGGG 1510

Qy   460 GlyAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisPro 479
Db  1511 GGTGACCATACTCAGCTGACAGATCTAAGTCCTGAAGAGAAGATGCTACCCAAACACCCC 1570

Qy   480 GluGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGly 499
Db  1571 GAAGGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAGAGAATCAAGGTACAGGGA 1630

Qy   500 SerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGln 519
Db  1631 AGTCCCCTGAAGAAGCTCTTCAGCAGTTCGGGCTTAAAGAAGCTCTCCGGGAAGAAGCAG 1690

Qy   520 LysGlyLysArg---GlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHis 538
Db  1691 AAGGGGAAGAGAGGAGGAGGCGGGGGAGATGAAGAGCCAGGAGAATACCAACACATTCAA 1750

Qy   539 ThrGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGlu 558
Db  1751 ACCGAGTCCCCAGAGAGTGCTGACGAGCAGAAGGGAGAGAGCTCTGCCTCTTCCCCTGAA 1810

Qy   559 GluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGlu 578
Db  1811 GAGCCCGAGGAGATCGCGTGTCTGGAGAAGGGGCCATCGGAAGCACCCCAGGAAGCGGAA 1870

Qy   579 AlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAla 598
Db  1871 GCTGAGGAAGGAGCGACTTCCGACGGAGAGAAGAAAAGGGAAGGGATCACCCCCTGGGCA 1930

Qy   599 SerPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLys 618
Db  1931 TCCTTCAAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGCGACAAG 1990

Qy   619 GluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSer 638
Db  1991 GAAGAAGAGCTGGATAAGGTCAAGAGTGCCACCTTGTCCTCCACGGAGAGCACGGCGTCT 2050

Qy   639 GluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArg 658
Db  2051 GGAATGCAGGATGAGGTCAGAGCGGTTGGCGAGGAGCAAAGGTCAGAGGAGCCAAAGCGC 2110

Qy   659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
Db  2111 AGGGTGGATACTTCAGTGTCTTGGGAGGCGTTGATTTGTGTCGGATCGTCCAAGAAGAGA 2170

Qy   679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
Db  2171 GCGAGGAAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGAACACTGGGAGGGGATGGC 2230

Qy   699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
Db  2231 CACAGAGCGGAGGAGGCTAGCAAAGACAAAGAAGCA------GATGCTCTTCCTGCCAGC 2284

Qy   719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
Db  2285 ACCCAGGAACAAGACCAAGCGCACGGAAGTTCCTCACCCGAGCCAGCTGGAAGCCCTTCT 2344

Qy   739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLys 758
Db  2345 GAAGGGGAGGGCGTCTCCACCTGGGAGTCATTTAAGAGATTAGTCACTCCACGAAAAAAA 2404

Qy   759 SerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGlu 778
Db  2405 TCCAAGTCAAAACTGGAAGAGAGAGCCGAAGACTCCGGTGCAGAGCAGTTGGCCTCCGAG 2464

Qy   779 IleGluProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArg 798
Db  2465 ATCGAACCAAGTAGAGAGGAATCTTGGGTTTCCATTAAGAAATTTATTCCTGGACGGCGG 2524

Qy   799 LysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGlu 818
Db  2525 AAGAAAAGGGCAGATGGAAGCAAGAACAGGCCGCCGTTGAAGACTCGGGGCCAGGAGAG 2584

Qy   819 IleAsnGluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaVal 838
Db  2585 ATCAATGAGGACGACCCCGACGTCCCAGCTGTTGTGCCTCTGTCTGAGTACGATGCGGTA 2644

Qy   839 GluArgGluLysMetGlu---------AlaGlnGlyAsnThrGluLeuProGlnLeuLeu 855
Db  2645 GAGAGAGAGAAGCTGGAAGCGCAGCGAGCTCAGGAGAACGTGGAGCTGCCCCAGCTGAAG 2704

Qy   856 GlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrValSerValAla 875
Db  2705 GGGGCTGTGTATGTGTCTGAGGAGCTTAGTAAGACTCTGGTTCACACTGTGAGTGTCGCG 2764

Qy   876 ValIleAspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSer 895
Db  2765 GTCATTGATGGGACCAGGGCAGTCACCAGTGCCGAAGAGCGGTCCCCTTCGTGGATATCT 2824

Qy   896 AlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGluGlu 915
Db  2825 GCTTCCATGACAGAACCTCTTGAGCACGCAGAGGGAGTGGCCACACCGCCTGTTGGAGAG 2884

Qy   916 ValThrGluLysAspIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGlu 935
Db  2885 GTCACTGAAAAAGACATCACTGCAGAAGCAACTCCTGCACTCGCCCAGACTTTACCAGGG 2944

Qy   936 GlyLysAspAlaHisAspAspMetValThrSerGluValAspPheThrSerGluAlaVal 955
Db  2945 GGCAAAGATGCCCATGACGACATAGTCACCAGTGAAGTGGATTTTACCTCAGAAGCAGTG 3004

Qy   956 ThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAla 975
Db  3005 ACAGCCGCAGAAACCACAGAGGCGCTCCGCGCTGAAGAACTTACCGAAGCATCAGGGGCA 3064
```

```
Qy     976 GluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThr 995
Db    3065 GAAGAGACCACAGACATGGTGTCTGCAGTTTCCCAGCTGTCCGACTCCCCGGACACCACA 3124

Qy     996 GluGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGlu 1015
Db    3125 GAGGAAGCCACCCCAGTTCAGGAGGTAGAGGGTGGCATGCTAGATACGGAAGAACAGGAG 3184

Qy    1016 ArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnVal 1035
Db    3185 CGCCAGACGCAGGCCGTCCTCCAAGCCGTTGCAGACAAAGTGAAAGAGGACTCCCAGGTG 3244

Qy    1036 ProAlaThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluVal 1055
Db    3245 CCTGCAACCCAGACTCTGCAGAGAGCAGGGCCGAAAGCACTGGAGAAGGTGGAGGAGGTA 3304

Qy    1056 GluGluAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyPro 1075
Db    3305 GAGGAGGACTCCGAGGTGCTGGCTACCGAGAAAGAGAAGGATGTTGTGCCGGAAGGACCC 3364

Qy    1076 ValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrPro 1095
Db    3365 GTGCAGGAAGCTGAAACTGAGCATCTTGCACAGGGCTCCGAGACTGTACAGGCTACCCCA 3424

Qy    1096 GluSerLeuGluValProGluValThrAlaAspValAspHisValAlaThrCysGlnVal 1115
Db    3425 GAGAGCCTTGAAGTTCCTGAAGTCACAGAGGATGTAGACCGTGCCACCACATGCCAGGTT 3484

Qy    1116 IleLysLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThr 1135
Db    3485 ATCAAGCACCAGCAGCTGATGGAACAGGCTGTGGCCCCTGAGTCATCTGAAACCTTGACA 3544

Qy    1136 AspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGln 1155
Db    3545 GACAGTGAGACAAATGGAAGTACTCCCCTCGCAGATTCAGACACTCCAAACGGGACACAG 3604

Qy    1156 GlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGln 1175
Db    3605 CAAGACGAGACCGTTGACAGCCAGGACAGTAATGCCATTGCCGCCGTCAAGCAGTCACAG 3664

Qy    1176 ValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsn 1195
Db    3665 GTCACTGAAGAGGAGGCAGCTGCTGCTCAGACGGAGGGGCCTTCAACA---CCATCTAGT 3721

Qy    1196 ValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProThrGln 1215
Db    3722 TTTCCAGCCCAGGAAGAACACAGGGAAAAACCAGGAAGGGATGTTCTAGAACCCACA--- 3778

Qy    1216 GlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGly 1235
Db    3779 CAAGCGCTGGCTGCCGGGGCAGTGCCTATTCTGGCAAAGGCTGAGGTGGGTCAAGAGGGT 3838

Qy    1236 GluValAspTrpLeuAspGlyGluLysValLysGlu---------------GluGlnGlu 1250
Db    3839 GAGGCTGGCCAGTTTGATGGAGAAAAAGTCAAAGACGGACAGTGTGTTAAAGAACTGGAG 3898

Qy    1251 ValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGlu 1270
Db    3899 GTGCCTGTGCACACTGGACCCAACAGTCAAAAGACTGCTGACTTGACACGTGACAGTGAA 3958

Qy    1271 ValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSerLeuSerLeu 1290
Db    3959 GTAATGGAAGTGGCCAGATGTCAGGAAACTGAGAGTAATGAAGAACAGAGTATTAGCCCG 4018

Qy    1291 GluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGluGln 1310
Db    4019 GAGAAAAGAGAGATGGGAACCGACGTTGAAAAGGAGGAAACAGAGACCAAGACAGAGCAA 4078

Qy    1311 ValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLys 1330
Db    4079 GCCAGTGAAGAACATGAGCAGGAAACAGCTGCTCCTGAGCATGAAGGAACCCACCCTAAG 4138

Qy    1331 ProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGly 1350
Db    4139 CCAGTCCTGACAGCTGACATGCCTCACTCAGAGAGGGGAAAGGCACTGGGCAGCCTTGAA 4198

Qy    1351 GlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSer 1370
Db    4199 GGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGATTGCATAGAGGTTCAAGTTCAAAGC 4258

Qy    1371 LeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrValVal 1390
Db    4259 TCAGACACACCAGTCACTCAAACAACCGAAGCTGTGAAAAAGGTCGAAGAAACTGTGGCA 4318

Qy    1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db    4319 ACTTCAGAGATGGATGAAAGTTTGGAGTGTGCAGGTGCGCAATCATTACCAGCTGAGAAG 4378

Qy    1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db    4379 CTCTCCGAAACCGGTGGCTACGGGACTCTTCAGCATGGAGAGGACACCGTGCCCCAGGGG 4438

Qy    1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db    4439 CCTGAGTCTCAGGCAGAGTCCATCCCAATAATAGTAACTCCTGCTCCTGAAAGCATCCTA 4498

Qy    1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db    4499 CATTCTGACCTTCAAAGAGAAGTGAGCGCATCCCAGAAACAGAGATCAGATGAAGATAAC 4558

Qy    1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db    4559 AAGCCAGATGCTGGTCCTGATGCTGCCGGCAAGGAGAGTGCAGCAAGAGAGAAAATCCTC 4618

Qy    1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
Db    4619 AGGGCTGAACCTGAGATCTTGGAACTTGAGAGTAAGAGCAATAAGATTGTCCAGAGTGTC 4678

Qy    1511 IleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaProGluThrHisAlaTyr 1530
Db    4679 ATCCAGACAGCCGTCGACCAGTTTGCACGTACAGAAACAGCCCCCGAAACCCACGCTTCT 4738

Qy    1531 AspSerGlnThrGlnValProAla-CysArgLeuAspSerArgGluProAsnArgCysTr 1550
Db    4739 GATTTACAGAATCAGGTTCCTGTGATGCAGGCTGACAGCCAGGGAGCACAACAGATGCTG 4798

Qy    1550 pThrLysMetLys---------------AspAlaLysMetLysHisProValProGlnPr 1565
Db    4799 GACAAAGATGAAAGCGACCTTCAAGTCTCCCCCAAGATGGAACACTCAGTGCCGTAGCC 4858

Qy    1565 oArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLe 1585
Db    4859 CAGGAAGGACTTGCGGTTTCTGATAGTTCTGAAGGCATGAGCAAGGCTTCAGAAATGATC 4918

Qy    1585 uProArgLeuGlnLeuLysAlaProValSerLys 1596
Db    4919 ACCACGCTTGCAGTTGAAAGTGCCAGTGTCAAAG 4952
```

RESULT 6
AF326230

```
LOCUS       AF326230                5902 bp    mRNA    linear   ROD 16-JUL-2001
DEFINITION  Mus musculus testis-specific SSeCKS mRNA, complete cds.
ACCESSION   AF326230
VERSION     AF326230.1  GI:13236407
KEYWORDS    .
SOURCE      Mus musculus.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 5902)
  AUTHORS   Camus,A., Mesbah,K., Rallu,M., Babinet,C. and Barra,J.
  TITLE     Gene trap insertion reveals two open reading frames in the mouse
            SSeCKS gene: the form predominantly detected in the nervous system
            is suppressed by the insertion while the other, specific of the
            testis, remains expressed
  JOURNAL   Mech. Dev. 105 (1-2), 79-91 (2001)
```

```
MEDLINE     21322683
  PUBMED    11429284
REFERENCE   2  (bases 1 to 5902)
  AUTHORS   Mesbah,K., Babinet,C. and Barra,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2000) Immunologie, Institut Pasteur, 25 rue du
            Docteur Roux, Paris 75015, France
FEATURES             Location/Qualifiers
     source          1..5902
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /chromosome="10"
                     /tissue_type="testis"
     CDS             160..4899
                     /codon_start=1
                     /product="testis-specific SSeCKS"
                     /protein_id="AAK16152.1"
                     /db_xref="GI:13236408"
                     /translation="MAASSTVVEDITKDEQEETPEIIEQIPASESNVEEMAQAAESQA
                     NDVGFKKVFKFVGFKFTVKKDKNEKSDTVQLLTVKKDEGEGAEASVGAGDHQEPGVET
                     VGESASKESELKQSTEKQEGTLKQAQSSTEIPLQAESGQGTEEEAAKDGEENREKEPT
                     KPLESPTSPVSNETTSSFKKFFTHGWAGWRKKTSFKKPKEDDLETSEKRKEQEAEKVD
                     EEEGEKTEPAPAEEQEPAEGTDQARLSADYEKVELPLEDQVGDLEALSEKCAPLATEV
                     FDEKTEAHQEVVAEVHVSTVEKMTKGQGGAEVEGDVVVEGSGESLPPEKLAETQEVPQ
                     EAEPVEELMKTKEVCVSGGDHTQLTDLSPEEKMLPKHPEGIVSEVEMLSSQERIKVQG
                     SPLKKLFSSSGLKKLSGKKQKGKRGGGGGDEEPGEYQHIQTESPESADEQKGESSASS
                     PEEPEEIACLEKGPSEAPQEAEAEEGATSDGEKKREGITPWASFKKMVTPKKRVRRPS
                     ESDKEEELDKVKSATLSSTESTASGMQDEVRAVGEEQRSEEPKRRVDTSVSWEALICV
                     GSSKKRARKASSSDDEGGPRTLGGDGHRAEEASKDKEADALPASTQEQDQAHGSSSPE
                     PAGSPSEGEGVSTWESFKRLVTPRKKSKSKLEERAEDSGAEQLASEIEPSREESWVSI
                     KKFIPGRRKKRADGKQEQAAVEDSGPGEINEDDPDVPAVVPLSEYDAVEREKLEAQRA
                     QENVELPQLKGAVYVSEELSKTLVHTVSVAVIDGTRAVTSAEERSPSWISASMTEPLE
                     HAEGVATPPVGEVTEKDITAEATPALAQTLPGGKDAHDDIVTSEVDFTSEAVTAAETT
                     EALRAEELTEASGAEETTDMVSAVSQLSDSPDTTEEATPVQEVEGGMLDTEEQERQTQ
                     AVLQAVADKVKEDSQVPATQTLQRAGPKALEKVEEVEEDSEVLATEKEKDVVPEGPVQ
                     EAETEHLAQGSETVQATPESLEVPEVTEDVDRATTCQVIKHQQLMEQAVAPESSETLT
                     DSETNGSTPLADSDTPNGTQQDETVDSQDSNAIAAVKQSQVTEEEAAAAQTEGPSTPS
                     SFPAQEEHREKPGRDVLEPTQALAAGAVPILAKAEVGQEGEAGQFDGEKVKDGQCVKE
                     LEVPVHTGPNSQKTADLTRDSEVMEVARCQETESNEEQSISPEKREMGTDVEKEETET
                     KTEQASEEHEQETAAPEHEGTHPKPVLTADMPHSERGKALGSLEGSPSLPDQDKADCI
                     EVQVQSSDTPVTQTTEAVKKVEETVATSEMDESLECAGAQSLPAEKLSETGGYGTLQH
                     GEDTVPQGPESQAESIPIIVTPAPESILHSDLQREVSASQKQRSDEDNKPDAGPDAAG
                     KESAAREKILRAEPEILELESKSNKIVQSVIQTAVDQFARTETAPETHASDLQNQVPV
                     MQADSQGAQQMLDKDESDLQVSPQDGTLSAVAQEGLAVSDSSEGMSKASEMITTLAVE
                     SASVKESVEKLPLQCKDEKEHAADGPQHQSLAKAEADASGNLTKESPDTNGPKLTEEG
                     DALKEEMNKAQTEEDDLQEPKGDLTES"
BASE COUNT     1774 a   1351 c   1677 g   1100 t
ORIGIN

Alignment Scores:
Pred. No.:              2.61e-215     Length:          5902
Score:                  6255.00       Matches:         1276
Percent Similarity:     88.58%        Conservative:    73
Best Local Similarity:  83.78%        Mismatches:      151
Query Match:            77.48%        Indels:          23
DB:                     10            Gaps:            11

US-09-902-432-4 (1-1596) x AF326230 (1-5902)

Qy       90 GlyGlnArgGluSerGluAspValArgGluLysAspArgValGluGluMetAlaAlaAsn 109
            ||||||||||||||||||||||||:::||||||||||    :::||||||||||||:::
Db      112 GGACAGAGAGAGTCAGAAGATGTGAAAGAAAAAGACCGAGCTAAAGAAATGGCAGCCAGT 171

Qy      110 SerThrAlaValGluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIleGlu 129
            |||||   ||||||||||||||||||||   |||||||||||   |||||||||||
Db      172 TCCACAGTTGTTGAAGATATCACAAAGGACGAGCAGGAGGAAACACCGGAAATAATCGAA 231

Qy      130 GlnIleProAlaSerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAla 149
            ||||||||||||||||||:::||||||||||||||   |||   ||||||||||||||
Db      232 CAGATCCCTGCTTCAGAGAGCAATGTGGAAGAAATGGCGCAGGCTGCTGAGTCCCAAGCT 291

Qy      150 AsnAspValGlyPheLysLysValPheLysPheValGlyPheLysPheThrValLysLys 169
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      292 AATGACGTCGGCTTCAAGAAGGTATTTAAATTTGTTGGTTTTAAATTCACGGTGAAGAAG 351

Qy      170 AspLysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGlu 189
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      352 GATAAAAACGAAAAGTCAGATACCGTCCAGCTACTCACTGTCAAGAAGGATGAAGGCGAA 411

Qy      190 GlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaVal 209
            |||||||||||||||||||||||||||||||||||||||||   ||||||||   |||
Db      412 GGGGCAGAAGCCTCCGTCGGAGCAGGAGACCACCAAGAGCCCGGAGTGGAGACC---GTC 468

Qy      210 GlyGluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThr 229
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      469 GGCGAATCAGCATCCAAAGAAAGTGAGCTGAAGCAATCCACAGAGAAGCAAGAAGGCACC 528

Qy      230 LeuLysGlnGluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAla 249
            ||||||||   ||||||||||||||||||||||||||||||||||||     |||
Db      529 CTGAAGCAAGCACAGAGCAGCACAGAAATTCCCCTTCAAGCCGAATCTGGTCAAGGGACC 588

Qy      250 GluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerPro 269
            |||||||||||      :::||||||||    :::||||||||||||||||||
Db      589 GAGGAAGAAGCAGCCAAAGATGGAGAAGAAAACCGAGAGAAAGAACCTACCAAGCCCCTA 648

Qy      270 GluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThr 289
            |||||||:::||||||||||:::::|||||||||||||||||||||||||||||||||||
Db      649 GAATCTCCGACCAGCCCTGTCAGCAATGAGACAACATCTTCCTTCAAGAAATTCTTCACT 708

Qy      290 HisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeu 309
            ||||||||||||||||||||||||||||||||||||||||||   ||||||||||||||
Db      709 CACGGGTGGGCCGGCTGGCGCAAGAAGACCAGCTTCAAGAAACCAAAGGAAGATGATCTG 768

Qy      310 GluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGlu 329
            ||||||:::|||||||||||||||||||||||||||||||||||||||||||    |||
Db      769 GAAACTTCCGAGAAGAGAAAGGAGCAAGAGGCTGAAAAAGTAGACGAGGAAGAAGGGGAA 828

Qy      330 LysThrGluProAla---SerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
            ||||||||||||||      :::||||||||||||||||||||    ||||||||||||||
Db      829 AAGACAGAGCCAGCCCCAGCCGAGGAGCAGGAGCCTGCAGAAGGCACAGACCAGGCCAGG 888

Qy      349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      889 TTGTCAGCCGACTATGAGAAGGTGGAGTTGCCTTTGGAAGACCAGGTCGGTGACCTGGAG 948

Qy      369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
            |||    |||   |||||||||||||||||||||||||||||||||||||||||    |||
Db      949 GCATTGTCG---GAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGACGGAA 1005

Qy      389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
            |||||||||||||||||||||||||||||||||||||||||||||||||      :::
Db     1006 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGATGACGAAAGGG 1065

Qy      409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
            |||   ||||||    |||    ||||||   |||||||||||.||||:::|||||||||||
Db     1066 CAA---GGAGGAGCAGAGGTGGAAGGGGATGTGGTGGTGGAAGGATCGGGAGAATCCTTG 1122

Qy      429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
            ||||||||||||||||||||||||   ||||||||||||||||||||||||||    ||||||
Db     1123 CCCCCTGAGAAACTGGCTGAGACCCAGGAGGTCCCCCAGGAAGCTGAGCCTGTGGAGGAG 1182

Qy      449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
            |||||||||::::::|||:::||||||||||||||||||||||||||||||||||||||||
Db     1183 CTGATGAAGACCAAAGAAGTATGCGTCTCTGGGGGTGACCATACTCAGCTGACAGATCTA 1242

Qy      469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
            ||||||||||||||||||     ||||||||||||||||||||||||||||||||||||
Db     1243 AGTCCTGAAGAGAAGATGCTACCCAAACACCCCGAAGGCATTGTCAGTGAGGTGGAGATG 1302

Qy      489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1303 CTGTCCTCTCAGGAGAGAATCAAGGTACAGGGAAGTCCCCTGAAGAAGCTCTTCAGCAGT 1362

Qy      509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArg---GlyGlyGlyGly 527
            |||||||||||||||||||||||||||||||||||||||||||||      ||||||||||
Db     1363 TCGGGCTTAAAGAAGCTCTCCGGGAAGAAGCAGAAGGGGAAGAGAGGAGGAGGCGGGGGA 1422
```

```
Qy    528 AspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGlu 547
Db   1423 GATGAAGAGCCAGGAGAATACCAACACATTCAAACCGAGTCCCCAGAGAGTGCTGACGAG 1482

Qy    548 GlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGlu 567
Db   1483 CAGAAGGGAGAGAGCTCTGCCTCTTCCCCTGAAGAGCCCGAGGAGATCGCGTGTCTGGAG 1542

Qy    568 LysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGly 587
Db   1543 AAGGGGCCATCGGAAGCACCCCAGGAAGCGGAAGCTGAGGAAGGAGCGACTTCCGACGGA 1602

Qy    588 GluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLys 607
Db   1603 GAGAAGAAAAGGGAAGGGATCACCCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAG 1662

Qy    608 LysArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSer 627
Db   1663 AAACGGGTCCGAAGACCTTCTGAGAGCGACAAGGAAGAAGAGCTGGATAAGGTCAAGAGT 1722

Qy    628 AlaThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrVal 647
Db   1723 GCCACCTTGTCCTCCACGGAGAGCACGGCGTCTGGAATGCAGGATGAGGTCAGAGCGGTT 1782

Qy    648 GlyGluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGlu 667
Db   1783 GGCGAGGAGCAAAGGTCAGAGGAGCCAAAGCGCAGGGTGGATACTTCAGTGTCTTGGGAG 1842

Qy    668 AlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAsp 687
Db   1843 GCGTTGATTTGTGTCGGATCGTCCAAGAAGAGAGCGAGGAAGGCATCCTCTTCAGATGAT 1902

Qy    688 GluGlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAsp 707
Db   1903 GAAGGAGGGCCAAGAACACTGGGAGGGGATGGCCACAGAGCGGAGGAGGCTAGCAAAGAC 1962

Qy    708 LysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGly 727
Db   1963 AAAGAAGCA------GATGCTCTTCCTGCCAGCACCCAGGAACAAGACCAAGCGCACGGA 2016

Qy    728 SerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGlu 747
Db   2017 AGTTCCTCACCCGAGCCAGCTGGAAGCCCTTCTGAAGGGGAGGGCGTCTCCACCTGGGAG 2076

Qy    748 SerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla 767
Db   2077 TCATTTAAGAGATTAGTCACTCCACGAAAAAAATCCAAGTCAAAACTGGAAGAGAGAGCC 2136

Qy    768 GluAspSerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrp 787
Db   2137 GAAGACTCCGGTGCAGAGCAGTTGGCCTCCGAGATCGAACCAAGTAGAGAGGAATCTTGG 2196

Qy    788 ValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGlu 807
Db   2197 GTTTCCATTAAGAAATTTATTCCTGGACGGCGGAAGAAAAGGGCAGATGGGAAGCAAGAA 2256

Qy    808 GlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValPro 827
Db   2257 CAGGCCGCCGTTGAAGACTCGGGGCCAGGAGAGATCAATGAGGACGACCCCGACGTCCCA 2316

Qy    828 AlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu--------- 844
Db   2317 GCTGTTGTGCCTCTGTCTGAGTACGATGCGGTAGAGAGAGAGAAGCTGGAAGCGCAGCGA 2376

Qy    845 AlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeu 864
Db   2377 GCTCAGGAGAACGTGGAGCTGCCCCAGCTGAAGGGGGCTGTGTATGTGTCTGAGGAGCTT 2436

Qy    865 SerLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThr 884
Db   2437 AGTAAGACTCTGGTTCACACTGTGAGTGTCGCGGTCATTGATGGGACCAGGGCAGTCACC 2496

Qy    885 SerValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHis 904
Db   2497 AGTGCCGAAGAGCGGTCCCCTTCGTGGATATCTGCTTCCATGACAGAACCTCTTGAGCAC 2556

Qy    905 ThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGlu 924
Db   2557 GCAGAGGGAGTGGCCACACCGCCTGTTGGAGAGGTCACTGAAAAAGACATCACTGCAGAA 2616

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   2617 GCAACTCCTGCACTCGCCCAGACTTTACCAGGGGGCAAAGATGCCCATGACGACATAGTC 2676

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   2677 ACCAGTGAAGTGGATTTTACCTCAGAAGCAGTGACAGCCGCAGAAACCACAGAGGCGCTC 2736

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   2737 CGCGCTGAAGAACTTACCGAAGCATCAGGGGCAGAAGAGACCACAGACATGGTGTCTGCA 2796

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   2797 GTTTCCCAGCTGTCCGACTCCCCGGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTA 2856

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   2857 GAGGGTGGCATGCTAGATACGGAAGAACAGGAGCGCCAGACGCAGGCCGTCCTCCAAGCC 2916

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThr 1044
Db   2917 GTTGCAGACAAAGTGAAAGAGGACTCCCAGGTGCCTGCAACCCAGACTCTGCAGAGAGCA 2976

Qy   1045 GlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSer 1064
Db   2977 GGGCCGAAAGCACTGGAGAAGGTGGAGGAGGTAGAGGAGGACTCCGAGGTGCTGGCTACC 3036

Qy   1065 GluLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeu 1084
Db   3037 GAGAAAGAGAAGGATGTTGTGCCGGAAGGACCCGTGCAGGAAGCTGAAACTGAGCATCTT 3096

Qy   1085 AlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThr 1104
Db   3097 GCACAGGGCTCCGAGACTGTACAGGCTACCCCAGAGAGCCTTGAAGTTCCTGAAGTCACA 3156

Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGln 1124
Db   3157 GAGGATGTAGACCGTGCCACCACATGCCAGGTTATCAAGCACCAGCAGCTGATGGAACAG 3216

Qy   1125 AlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrPro 1144
Db   3217 GCTGTGGCCCCTGAGTCATCTGAAACCTTGACAGACAGTGAGACAAATGGAAGTACTCCC 3276

Qy   1145 LeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAsp 1164
Db   3277 CTCGCAGATTCAGACACTCCAAACGGGACACAGCAAGACGAGACCGTTGACAGCCAGGAC 3336

Qy   1165 SerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAla 1184
Db   3337 AGTAATGCCATTGCCGCCGTCAAGCAGTCACAGGTCACTGAAGAGGAGGCAGCTGCTGCT 3396

Qy   1185 GlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGlu 1204
Db   3397 CAGACGGAGGGGCCTTCAACA---CCATCTAGTTTTCCAGCCCAGGAAGAACACAGGGAA 3453

Qy   1205 GluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValPro 1224
Db   3454 AAACCAGGAAGGGATGTTCTAGAACCCACA---CAAGCGCTGGCTGCCGGGGCAGTGCCT 3510

Qy   1225 ValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLys 1244
Db   3511 ATTCTGGCAAAGGCTGAGGTGGGTCAAGAGGGTGAGGCTGGCCAGTTTGATGGAGAAAAA 3570

Qy   1245 ValLysGlu--------------GluGlnGluValPheValHisSerGlyProAsnSer 1259
```

```
Db   3571 GTCAAAGACGGACAGTGTGTTAAAGAACTGGAGGTGCCTGTGCACACTGGACCCAACAGT 3630

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   3631 CAAAAGACTGCTGACTTGACACGTGACAGTGAAGTAATGGAAGTGGCCAGATGTCAGGAA 3690

Qy   1280 LysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspVal 1299
Db   3691 ACTGAGAGTAATGAAGAACAGAGTATTAGCCCGGAGAAAAGAGAGATGGAACCGACGTT 3750

Qy   1300 GluLysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThr 1319
Db   3751 GAAAAGGAGGAAACAGAGACCAAGACAGAGCAAGCCAGTGAAGAACATGAGCAGGAAACA 3810

Qy   1320 AlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSer 1339
Db   3811 GCTGCTCCTGAGCATGAAGGAACCCACCCTAAGCCAGTCCTGACAGCTGACATGCCTCAC 3870

Qy   1340 SerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAsp 1359
Db   3871 TCAGAGAGGGGAAAGGCACTGGGCAGCCTTGAAGGAAGCCCTTCTCTCCCAGACCAAGAC 3930

Qy   1360 LysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAla 1379
Db   3931 AAAGCAGATTGCATAGAGGTTCAAGTTCAAAGCTCAGACACACCAGTCACTCAAACAACC 3990

Qy   1380 GluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGlu 1399
Db   3991 GAAGCTGTGAAAAAGGTCGAAGAAACTGTGGCAACTTCAGAGATGGATGAAAGTTTGGAG 4050

Qy   1400 CysValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThr 1419
Db   4051 TGTGCAGGTGCGCAATCATTACCAGCTGAGAAGCTCTCCGAAACCGGTGGCTACGGGACT 4110

Qy   1420 LeuGlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIlePro 1439
Db   4111 CTTCAGCATGGAGAGGACACCGTGCCCCAGGGGCCTGAGTCTCAGGCAGAGTCCATCCCA 4170

Qy   1440 IleIleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSer 1459
Db   4171 ATAATAGTAACTCCTGCTCCTGAAAGCATCCTACATTCTGACCTTCAAAGAGAAGTGAGC 4230

Qy   1460 AlaSerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAsp 1479
Db   4231 GCATCCCAGAAACAGAGATCAGATGAAGATAACAAGCCAGATGCTGGTCCTGATGCTGCC 4290

Qy   1480 GlyLysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeu 1499
Db   4291 GGCAAGGAGAGTGCAGCAAGAGAGAAAATCCTCAGGGCTGAACCTGAGATCTTGGAACTT 4350

Qy   1500 GluSerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAla 1519
Db   4351 GAGAGTAAGAGCAATAAGATTGTCCAGAGTGTCATCCAGACAGCCGTCGACCAGTTTGCA 4410

Qy   1520 ArgThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAla-Cy 1539
Db   4411 CGTACAGAAACAGCCCCCGAAACCCACGCTTCTGATTTACAGAATCAGGTTCCTGTGATG 4470

Qy   1539 sArgLeuAspSerArgGluProAsnArgCysTrpThrLysMetLys-------------- 1554
Db   4471 CAGGCTGACAGCCAGGGAGCACAACAGATGCTGGACAAAGATGAAAGCGACCTTCAAGTC 4530

Qy   1555 -AspAlaLysMetLysHisProValProGlnProArgGluAspLeuGlnValLeuThrVa 1574
Db   4531 TCCCCCCAAGATGGAACACTCAGTGCCGTAGCCCAGGAAGGACTTGCGGTTTCTGATAGT 4590

Qy   1574 lLeuGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVa 1594
Db   4591 TCTGAAGGCATGAGCAAGGCTTCAGAAATGATCACCACGCTTGCAGTTGAAAGTGCCAGT 4650

Qy   1594 lSerLys 1596
Db   4651 GTCAAAG 4657
```

```
RESULT 7
AR071282
LOCUS       AR071282                 5134 bp    DNA     linear   PAT 18-FEB-2000
DEFINITION  Sequence 1 from patent US 5910442.
ACCESSION   AR071282
VERSION     AR071282.1  GI:7222170
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 5134)
  AUTHORS   Gelman,I.H.
  TITLE     Tumor suppressor gene
  JOURNAL   Patent: US 5910442-A 1 08-JUN-1999;
FEATURES             Location/Qualifiers
     source          1. .5134
                     /organism="unknown"
BASE COUNT     1511 a   1167 c   1451 g   1005 t
ORIGIN

Alignment Scores:
Pred. No.:              1.95e-206      Length:        5134
Score:                  6005.00        Matches:       1232
Percent Similarity:     97.33%         Conservative:  7
Best Local Similarity:  96.78%         Mismatches:    27
Query Match:            74.38%         Indels:        14
DB:                     6              Gaps:          4

US-09-902-432-4 (1-1596) x AR071282 (1-5134)

Qy    329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
Db      2 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 61

Qy    349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
Db     62 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 121

Qy    369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
Db    122 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 181

Qy    389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
Db    182 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 241

Qy    409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
Db    242 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 301

Qy    429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
Db    302 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 361

Qy    449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
Db    362 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 421

Qy    469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
Db    422 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 481

Qy    489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
Db    482 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 541

Qy    509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
Db    542 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 601

Qy    529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
```

```
Db     602 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 661

Qy     549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568

Db     662 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 721

Qy     569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588

Db     722 GGGCCGCTGGAAGCACCCAG-GATGGGGAAGCTGAGGAAGGAACTACTTCGT--GGAGAG 778

Qy     589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608

Db     779 AAGAAGAGGAAGG--ATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 836

Qy     609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628

Db     837 CGGTC-CGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 895

Qy     629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648

Db     896 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 955

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668

Db     956 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 1015

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688

Db    1016 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATATAAGA 1075

Qy     689 GlyGlyProArgThrLeuGly-GlyAspSerHisArgAlaGluGluAlaSerLysAspLy 708

Db    1076 ---GGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAA 1132

Qy     708 sGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySe 728

Db    1133 AGAAGCCGA-ACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAG 1191

Qy     728 rSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSe 748

Db    1192 TTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTC 1251

Qy     748 rPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla-- 767

Db    1252 ATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA-AG 1310

Qy     768 ----GluAspSerSerVal---GluGlnLeuSerThrGluIleGluProSerArgGluGl 785

Db    1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy     785 uSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLy 805

Db    1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy     805 s-GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProA 825

Db    1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy     825 snValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluA 845

Db    1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy     845 laGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865

Db    1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy     865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885

Db    1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy     885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905

Db    1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy     905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluG 925

Db    1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy     925 luThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValT 945

Db    1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy     945 hrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuA 965

Db    1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy     965 rgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaV 985

Db    1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy     985 alSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValG 1005

Db    1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy    1005 luSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaV 1025

Db    2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy    1025 alAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrG 1045

Db    2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy    1045 lySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerG 1065

Db    2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy    1065 luLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuA 1085

Db    2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy    1085 laGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrA 1105

Db    2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy    1105 laAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnA 1125

Db    2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy    1125 laValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProL 1145

Db    2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy    1145 euAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspS 1165

Db    2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy    1165 erLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaG 1185

Db    2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy    1185 lnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluG 1205

Db    2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy    1205 luProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProV 1225

Db    2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy    1225 alLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysV 1245

Db    2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy    1245 alLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspV 1265

Db    2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807
```

```
Qy     1265 alThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluV 1285
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy     1285 alGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgG 1305
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy     1305 luThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisG 1325
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy     1325 luGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysA 1345
            ||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy     1345 laLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleG 1365
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy     1365 luValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysV 1385
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy     1385 alIleGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisL 1405
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy     1405 euLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluA 1425
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy     1425 spThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProA 1445
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy     1445 laProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluA 1465
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy     1465 rgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrA 1485
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy     1485 laIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnL 1505
            |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy     1505 ysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaP 1525
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy     1525 roGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgG 1545
            ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||||
Db     3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy     1545 luProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnP 1565
            |||||||||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db     3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy     1565 roArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysL 1585
            |||||||||||||||||||||||||||||||||||||||||||         ||||||||||
Db     3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGC--TCAGCTCGGAAATGCT 3765

Qy     1585 euProArgLeuGlnLeuLysAlaProValSerLys 1596
            ||||||||||||||||||||||||||||||||||||
Db     3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 3800

RESULT 8
AX405857
LOCUS       AX405857                 6614 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 272 from Patent WO0222660.
ACCESSION   AX405857
VERSION     AX405857.1  GI:21439153
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tang,Y.T., Liu,C., Zhou,P., Asundi,V., Zhang,J., Zhao,Q.A., Ren,F.,
            Xue,A.J., Yang,Y., Wehrman,T. and Drmanac,R.T.
  TITLE     Novel nucleic acids and polypeptides
  JOURNAL   Patent: WO 0222660-A 272 21-MAR-2002;
            HYSEQ, INC. (US)
FEATURES             Location/Qualifiers
     source          1..6614
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             192..5543
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD34964.1"
                     /db_xref="GI:21439154"
                     /translation="MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADP
                     AIAASDPATKLLQKNGQLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEV
                     IVTEVGQRDSEDVSERDSDKEMATKSAVVHDITDDGQEETPEIIEQIPSSESNLEELT
                     QPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVKKDEGEGAAGAGDHQDP
                     SLGAGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAVEECKEEGEEKQEKE
                     PSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEK
                     VDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPS
                     EEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTEVEETAGSVPAEELVEM
                     DAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQER
                     MKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGDEESGEHTQVPADSPDSQEEQKGESS
                     ASSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVR
                     RPSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEAL
                     ICVGSSKKRARRSSSDEEGGPKAMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQG
                     SSSPEQAGSPTEGEGVSTWESFKRLVTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEP
                     GKEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVE
                     REKMEAQQAQKGAEQPEQKAATEVSKELSESQVHMMAAAVADGTRAATIIEERSPSWI
                     SASVTEPLEQVEAEAALLTEEVLEREVIAEEEPPTVTEPLPENREARGDTVVSEAELT
                     PEAVTAAETAGPLGSEEGTEASAAEETTEMVSAVSQLTDSPDTTEEATPVQEVEGGVP
                     DIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDVLQPVQRAEAERPEEQAEASGLKK
                     ETDVVLKVDAQEAKTEPFTQGKVVGQTTPESFEKAPQVTESIESSELVTTCQAETLAG
                     VKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIHEENEVASGT
                     QSGGTEAEAVPAQKERPPAPSSFVFQEETKEQSKMEDTLEHTDKEVSVETVSILSKTE
                     GTQEADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDAL
                     ELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSKQLLQTV
                     NVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLGETAN
                     ILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATTKK
                     GLSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEEDLEPENGILELETKSSKLVQ
                     NIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQETEKEGEDPQASAQDET
                     PITSAKEESESTAVGQAHSDISKDMSEASEKTMTVEVEGSTVNDQQLEEVVLPSEEEG
                     GGAGTKSVPEDDGHALLAERIEKSLVEPKEDEKGDDVDDPENQNSALADTDASGGLTK
                     ESPDTNGPKQKEKEDAQEVELQEGKVHSESDKAITPQAQEELQKQERESAKSELTES"
BASE COUNT     2094 a   1420 c   1827 g   1273 t
ORIGIN

Alignment Scores:
Pred. No.:              6.91e-147       Length:         6614
Score:                  4340.50         Matches:        970
Percent Similarity:     69.13%          Conservative:   190
Best Local Similarity:  57.81%          Mismatches:     408
Query Match:            53.77%          Indels:         110
DB:                     6               Gaps:           33

US-09-902-432-4 (1-1596) x AX405857 (1-6614)

Qy        1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
            ||||||||||||||||||||||||||||||||||||||||||    |||   ||||||
Db      192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGAGCTCC 251

Qy       20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
            ||||||:::|||           ||||||    ||||||:::|||||||||:::
Db      252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311
```

```
Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln----------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACAGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGACACCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 671

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    672 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 731

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    732 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 791

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    792 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 842

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    843 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 896

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    897 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 956

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    957 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1013

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1014 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1073

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1074 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1133

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1134 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1193

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1194 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1253

Qy    338 -----------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1254 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1313

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1314 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1373

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1374 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1433

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1434 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1484

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1485 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1529

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1530 TTGGTTGAAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1589

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1590 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1649

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1650 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1709

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1710 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1769

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1770 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1826

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1827 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1886

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1887 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1946

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1947 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2006

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2007 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2066

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2067 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2126

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2127 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2186

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2187 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2246

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2247 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2306

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2307 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2366

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2367 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGAAGTTCCTCCCCGGAG 2426

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
```

```
Db   2427 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2486

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769

Db   2487 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2546

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788

Db   2547 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2606

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808

Db   2607 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2666

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828

Db   2667 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2726

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845

Db   2727 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2786

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865

Db   2787 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2846

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885

Db   2847 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2906

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905

Db   2907 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2966

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924

Db   2967 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3026

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944

Db   3027 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3086

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964

Db   3087 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3146

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984

Db   3147 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3206

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004

Db   3207 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3266

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024

Db   3267 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3326

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038

Db   3327 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3386

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057

Db   3387 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3428

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077

Db   3429 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3488

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097

Db   3489 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3548

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113

Db   3549 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3608

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127

Db   3609 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3668

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147

Db   3669 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3728

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167

Db   3729 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3788

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187

Db   3789 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3848

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207

Db   3849 AGGCCTCCAGCA---CCTTCCAGTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3905

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226

Db   3906 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3965

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246

Db   3966 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4019

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259

Db   4020 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4079

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279

Db   4080 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4139

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294

Db   4140 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4199

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314

Db   4200 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4259

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333

Db   4260 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4319

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352

Db   4320 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4379

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372

Db   4380 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4439

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390

Db   4440 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4499

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410

Db   4500 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4559

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430

Db   4560 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4619
```

```
Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
          |||:::   ||||||:::|||   |||:::||||||:::          :::      |||
Db   4620 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4679

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
            ||||||:::||||||   :::     |||    :::    :::|||:::|||    |||
Db   4680 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4739

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
          :::         |||       :::          |||   |||   |||||||||:::   |||
Db   4740 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTA 4793

Qy   1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
          :::   |||        ||||||||||||||:::||||||:::|||:::|||    |||:::
Db   4794 GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATC 4853

Qy   1511 IleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThrHisAla 1529
          ||||||||||||||||||||||||    ||||||    |||||||||    |||
Db   4854 ATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACG 4913

Qy   1530 TyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCys 1549
             :::   |||||||||           :::      ||||||::::::        :::
Db   4914 TCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAA 4973

Qy   1550 TrpThrLys--------------MetLysAspAla-LysMetLysHisProValProGl 1564
                |||                    |||  :::|||||||||    :::    ||
Db   4974 ACGGAGAAAGAAGGAGAGGACCCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCA 5033

Qy   1564 nProArgGlu--------------------AspLeuGlnValLeuThrValLeuGluAl 1577
          ||||:::                         |||      :::|||          :::
Db   5034 GCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGAC 5093

Qy   1577 aTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
                :::|||:::|||             |||:::    |||    |||:::
Db   5094 ATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5145
```

RESULT 9
AX329766

```
LOCUS       AX329766                6608 bp    DNA     linear   PAT 09-JAN-2002
DEFINITION  Sequence 275 from Patent WO0194629.
ACCESSION   AX329766
VERSION     AX329766.1  GI:18102744
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,
            Horrigan,S., Soppet,D.R. and Weaver,Z.
  TITLE     Cancer gene determination and therapeutic screening using signature
            gene sets
  JOURNAL   Patent: WO 0194629-A 275 13-DEC-2001;
            Avalon Pharmaceuticals (US)
FEATURES             Location/Qualifiers
     source          1. .6608
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT     2091 a   1418 c    1827 g    1272 t
ORIGIN

Alignment Scores:
Pred. No.:              1.85e-145       Length:         6608
Score:                  4300.50         Matches:        967
Percent Similarity:     68.81%          Conservative:   189
Best Local Similarity:  57.56%          Mismatches:     408
Query Match:            53.27%          Indels:         116
DB:                     6               Gaps:           36

US-09-902-432-4 (1-1596) x AX329766 (1-6608)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
          ||||||||||||||||||||||||||||||||||||||||||    |||    ||||||
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
          ||||||:::|||             ||||||   ||||||:::|||||||||:::
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
                |||||||||   |||       ||||||||||||||||||   ||||||||||||
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
          |||||||||::::::||||||||||||||||||    ::::::    :::||||||   :::
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln------------------------------------------GluGlu 82
                ||||||                                          ||||||
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
          ||||||:::         :::||||||||||||:::||||||||||||    |||:::|||
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
             :::|||||||||          |||       |||    |||||||||    ||||||||||||
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
          |||            ||||||||||||||||:::||||||:::|||:::||||||:::
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
          ||||||   |||||||||||||||||||:::||||||||||||||||||||||||||||||
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
          |||||||||||||||||||||||||||    ||||||   ||||||||||||||||||||||
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
          |||||||||||||||||||||||||||               ||||||||||||||||:::
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
          |||||||:::                ||||||:::||||||||||||||||   |||||||||
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
          |||||||||    |||    |||||||||:::|||||||||         ||||||
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
          |||||||||   ||||||          ||||||    |||:::|||||||||||||||||||||
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
          |||||||||:::||||||   ||||||||||:::|||||||||   |||||||||   |||
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
          :::||||||||||||||||||    |||||||||||||||||||||||||||||||||||:::
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
          |||   ||||||||||::::::|||      :::||||||:::||||||||||||   ||||||
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
          ||||||   ||||||      |||    |||    |||||||||:::
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
                      ||||||||||||||         :::   ||||||||||||:::|||
```

```
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310
Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370
Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430
Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481
Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526
Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586
Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646
Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706
Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766
Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823
Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883
Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943
Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003
Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063
Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123
Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183
Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243
Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303
Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363
```

```
Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423
Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483
Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543
Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603
Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663
Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723
Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783
Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843
Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903
Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963
Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023
Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083
Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143
Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203
Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263
Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323
Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383
Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA-----------------GAGGCAGAAAGACCAGAAGAG 3425
```

```
Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3845

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3846 AGGCCTCCAGCA---CCTTCCAGTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3902

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3903 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3962

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3963 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4016

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4017 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4076

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4077 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4136

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4137 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4196

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4197 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4256

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4257 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4316

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4317 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4376

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4377 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4436

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4437 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4496

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4497 ATTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4556

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4557 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4616

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4617 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4676

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4677 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4736

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4737 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4787

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
Db   4788 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4841

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
Db   4842 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4901

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
Db   4902 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4961

Qy   1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
Db   4962 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5021

Qy   1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
Db   5022 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5081

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   5082 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5139
```

```
RESULT 10
AX410793
LOCUS       AX410793                6608 bp    DNA     linear   PAT 14-JUN-2002
DEFINITION  Sequence 3440 from Patent WO0229103.
ACCESSION   AX410793
VERSION     AX410793.1  GI:21443498
KEYWORDS    .
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Alvares,C., Horne,D., Peres-da-Silva,S. and Vockley,J.G.
  TITLE     Gene expression profiles in liver cancer
  JOURNAL   Patent: WO 0229103-A 3440 11-APR-2002;
            GENE LOGIC INC (US)
FEATURES             Location/Qualifiers
     source          1..6608
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="EMBL/GenBank Accession No. U81607"
BASE COUNT     2091 a   1418 c   1827 g   1272 t
ORIGIN

Alignment Scores:
Pred. No.:              1.85e-145      Length:         6608
Score:                  4300.50        Matches:        967
Percent Similarity:     68.81%         Conservative:   189
Best Local Similarity:  57.56%         Mismatches:     408
Query Match:            53.27%         Indels:         116
```

```
DB:                    6            Gaps:        36

US-09-902-432-4 (1-1596) x AX410793 (1-6608)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln---------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGlnSerSerThrGluIleProLeuGln 242
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
```

```
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303
Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363
Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423
Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483
Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAGCGAAGACTCCATAGCT 2543
Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603
Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663
Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723
Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783
Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843
Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903
Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963
Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023
Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083
Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143
Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203
Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263
Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323
Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425
Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485
Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGCAGACCACCCCAGAAAGC 3545
Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605
Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665
Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725
Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785
Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3845
Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3846 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3902
Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3903 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3962
Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3963 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4016
Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4017 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4076
Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4077 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4136
Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4137 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4196
Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4197 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4256
Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4257 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4316
Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4317 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4376
Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4377 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4436
```

```
Qy    1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
              :::    |||   ||||||   |||   |||:::   :::      ||||||
Db    4437 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4496

Qy    1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuProAlaGluLys 1410
           |||   ||||||||||||:::   |||         |||||||||||:::      ||||||
Db    4497 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4556

Qy    1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
           ||||||                 :::                  ||||||   ||||||   |||
Db    4557 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4616

Qy    1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
           |||:::   ||||||:::|||   |||:::||||||:::         :::       |||
Db    4617 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4676

Qy    1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
                 ||||||:::||||||   :::   |||   :::   :::|||:::|||   |||
Db    4677 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4736

Qy    1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
           :::      |||      :::         |||   |||   |||||||||   :::
Db    4737 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4787

Qy    1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
                 |||||||||      |||||||||||||||:::||||||:::|||:::|||
Db    4788 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4841

Qy    1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
           |||:::||||||||||||||||||||||||   ||||||   |||||||||   |||
Db    4842 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4901

Qy    1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
                 :::   |||||||||         :::   ||||||::::::
Db    4902 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4961

Qy    1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
           :::         |||                   |||  :::|||||||||      ::
Db    4962 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5021

Qy    1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
           :     ||||||:::                    |||      :::|||
Db    5022 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5081

Qy    1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
             :::         :::|||:::|||         |||:::   |||   |||:::
Db    5082 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5139
```

RESULT 11
HSU81607

```
LOCUS       HSU81607                6608 bp    mRNA    linear   PRI 26-JUN-1997
DEFINITION  Homo sapiens gravin mRNA, complete cds.
ACCESSION   U81607
VERSION     U81607.1  GI:2218076
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6608)
  AUTHORS   Nauert,J.B., Klauck,T.M., Langeberg,L.K. and Scott,J.D.
  TITLE     Gravin, an autoantigen recognized by serum from myasthenia gravis
            patients, is a kinase scaffold protein
  JOURNAL   Curr. Biol. 7 (1), 52-62 (1997)
  MEDLINE   97153077
   PUBMED   9000000
REFERENCE   2  (bases 1 to 6608)
  AUTHORS   Nauert,J.B., Klauck,T.K., Langeberg,L.K. and Scott,J.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-DEC-1996) Biochemistry, OHSU, 3181 SW Sam Jackson
            Parkway, Portland, OR 97201, USA
COMMENT     On Jun 26, 1997 this sequence version replaced gi:1850362.
FEATURES             Location/Qualifiers
     source          1..6608
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     CDS             192..5537
                     /note="kinase scaffold protein"
                     /codon_start=1
                     /product="gravin"
                     /protein_id="AAC51366.1"
                     /db_xref="GI:2218077"
                     /translation="MGAGSSTEQRSPEQPPEGSSTPAEPEPSGGGPSAEAAPDTTADP
                     AIAASDPATKLLQKNGQLSTINGVAEQDELSLQEGDLNGQKGALNGQGALNSQEEEEV
                     IVTEVGQRDSEDVSERDSDKEMATKSAVVHDITDDGQEENRNIEQIPSSESNLEELTQ
                     PTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVKKDEGEGAAGAGDHQDPS
                     LGAGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQAVEECKEEGEEKQEKEP
                     SKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKEDEVEASEKKKEQEPEKV
                     DTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEKVELPSEEQVSGSQGPSE
                     EKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTEVEETAGSVPAEELVGMD
                     AEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKPPEGVVSEVEMLSSQERM
                     KVQGSPLKKLFTSTGLKKLSGKKQKGKRGGGDEESGEHTQVPADSPDSQEEQKGESSA
                     SSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVTPWASFKKMVTPKKRVRR
                     PSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEPKPEEPKRKVDTSVSWEALI
                     CVGSSKKRARRRSSSDEEGGPKAMGGDHQKADEAGKDKETGTDGILAGSQEHDPGQGS
                     SSPEQAGSPTEGEGVSTWESFKRLVTPRKKSKSKLEEKSEDSIAGSGVEHSTPDTEPG
                     KEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPTGANEDDSDVPAVVPLSEYDAVER
                     EKMEAQQAQKGAEQPEQKAATEVSKELSESQVHMMAAAVADGTRAATIIEERSPSWIS
                     ASVTEPLEQVEAEAALLTEEVLEREVIAEEEPPTVTEPLPENREARGDTVVSEAELTP
                     EAVTAAETAGPLGSEEGTEASAAEETTEMVSAVSQLTDSPDTTEEATPVQEVEGGVPD
                     IEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDVLQPVQRAEAERPEEQAEASGLKKE
                     TDVVLKVDAQEAKTEPFTQGKVVGQTTPESFEKAPQVTESIESSELVTTCQAETLAGV
                     KSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQKDEIVEIHEENEVASGTQ
                     SGGTEAEAVPAQKERPPAPSSFVFQEETKEQSKMEDTLEHTDKEVSVETVSILSKTEG
                     TQEADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVALKGEGTEEAECKKDDALE
                     LQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHETAVTVSEEVSKQLLQTVN
                     VPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFTLTAAAEEEKVLGETANI
                     LETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPDCQAKSTPVIVSATTKKG
                     LSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEDLEPENGILELETKSSKLVQNI
                     IQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQETEKEGEEPQASAQDETPI
                     TSAKEESESTAVGQAHSDISKDMSEASEKTMTVEVEGSTVNDQQLEEVVLPSEEEGGG
                     AGTKSVPEDDGHALLAERIEKSLVEPKEDEKGDDVDDPENQNSALADTDASGGLTKES
                     PDTNGPKQKEKEDAQEVELQEGKVHSESDKAITPQAQEELQKQERESAKSELTES"
BASE COUNT     2091 a   1418 c   1827 g   1272 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:              1.85e-145       Length:         6608
Score:                  4300.50         Matches:        967
Percent Similarity:     68.81%          Conservative:   189
Best Local Similarity:  57.56%          Mismatches:     408
Query Match:            53.27%          Indels:         116
DB:                     9               Gaps:           36
```

US-09-902-432-4 (1-1596) x HSU81607 (1-6608)

```
Qy       1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
           ||||||||||||||||||||||||||||||||||||||||||   |||   ||||||
Db     192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy      20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
           ||||||:::|||               ||||||   ||||||:::|||||||||:::
Db     252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy      39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
                 |||||||||   |||         |||||||||||||||||   ||||||||||||
Db     312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy      57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
           |||||||||::::::||||||||||||||||||      ::::::   :::||||||   :::
Db     372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy      77 GlnGluGlyGln------------------------------------------GluGlu 82
                 ||||||                                          ||||||
Db     432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491
```

```
Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
```

```
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3845

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3846 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3902

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3903 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3962

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3963 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4016

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4017 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4076

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4077 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4136

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4137 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4196

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4197 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4256

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4257 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4316

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4317 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4376

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4377 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4436

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4437 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4496

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4497 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4556

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4557 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4616

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4617 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4676

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4677 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4736

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4737 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4787
```

```
Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
                ||||||||      |||||||||||||:::|||||:::|||:::|||
Db   4788 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4841

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
          |||:::||||||||||||||||||||||||    ||||||   ||||||||   |||
Db   4842 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4901

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
                 :::  ||||||||          :::   ||||||::::::
Db   4902 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4961

Qy   1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
          :::       |||                   |||  :::||||||||    ::
Db   4962 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5021

Qy   1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
          :    ||||||:::                       |||       :::|||
Db   5022 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5081

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
           :::             :::|||:::|||           |||:::   |||  |||:::
Db   5082 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5139
```

RESULT 12
AR002603

```
LOCUS       AR002603                 6605 bp    DNA     linear   PAT 04-DEC-1998
DEFINITION  Sequence 4 from patent US 5741890.
ACCESSION   AR002603
VERSION     AR002603.1  GI:3964157
KEYWORDS    .
SOURCE      Unknown.
  ORGANISM  Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 6605)
  AUTHORS   Scott,J.D., Nauert,J.Brian. and Klauck,T.M.
  TITLE     Protein binding fragments of gravin
  JOURNAL   Patent: US 5741890-A 4 21-APR-1998;
FEATURES             Location/Qualifiers
     source          1..6605
                     /organism="unknown"
BASE COUNT     2091 a   1416 c   1826 g   1272 t
ORIGIN

Alignment Scores:
Pred. No.:               9.21e-145     Length:         6605
Score:                   4281.00       Matches:        965
Percent Similarity:      68.69%        Conservative:   189
Best Local Similarity:   57.44%        Mismatches:     409
Query Match:             53.03%        Indels:         117
DB:                      6             Gaps:           37

US-09-902-432-4 (1-1596) x AR002603 (1-6605)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
          |||||||||||||||||||||||||||||||||||||||||   |||  |||||
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
          ||||||:::|||          |||||  ||||||:::|||||||||:::
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
                |||||||||   |||          ||||||||||||||||   |||||||||||
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
          |||||||||::::::||||||||||||||||   ::::::   :::||||||    :::
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln-----------------------------------------GluGlu 82
                ||||||                                          ||||||
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
          ||||||:::        :::|||||||||||||::::||||||||||||    |||:::|||
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
             :::|||||||||         |||        |||   ||||||||  ||||||||||
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
          |||          ||||||||||||:::||||||:::|||:::||||||:::
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
          ||||||   |||||||||||||||||:::||||||||||||||||||||||||||||||||
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
          ||||||||||||||||||||||||      ||||||   ||||||||||||||||||||
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
          ||||||||||||||||||||||||||            ||||||||||||||||||:::
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
          ||||||:::            ||||||:::||||||||||||||||||    |||||||||
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
          ||||||||   |||   ||||||||:::|||||||||       ||||||
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
          |||||||||   ||||||      ||||||   |||:::||||||||||||||||||||||
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
          |||||||||:::||||||    ||||||||||:::|||||||||   |||||||||   |||
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
          :::|||||||||||||||||||   ||||||||||||||||||||||||||||||||:::
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
          |||   ||||||||||:::::::|||   :::||||||:::||||||||||||    ||||||
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
          ||||||   ||||||     |||   |||   ||||||||||:::
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
                      ||||||||||||||||          :::   ||||||||||||:::|||
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
          ||||||||||||||||||   |||:::||||||          :::     |||||||||
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
          |||   ||||||||||||||||||||||||||||||:::|||    ||||||   |||
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
          ||||||||||||||||||||||||||||||:::         |||||||||
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
                          ||||||   |||      |||:::|||   |||:::
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526
```

```
Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
```

```
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
             |||               ||||||:::||||||  ::::::      :::
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
                      |||         :::  ||||||   ||||||     |||||||||||
Db   3786 GTGCATCTGGTA---CCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3842

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
             |||          |||::::::         ||||||||     |||:::
Db   3843 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3899

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
                |||   ||||||  |||     :::|||::::::           |||    :::|||
Db   3900 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3959

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
          :::|||||||||         |||           |||   |||        ||||||    |||
Db   3960 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4013

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
                         |||     |||     :::   :::|||             :::
Db   4014 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4073

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
          :::|||      :::|||              |||         |||    |||::::::
Db   4074 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4133

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
             ::::::   |||:::||||||                      |||       |||   |||
Db   4134 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4193

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
          |||          |||||||:::||||||     |||    :::|||          |||:::||||||
Db   4194 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4253

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
                |||   ||||||||           |||         :::   :::
Db   4254 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4313

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
          |||:::::::::|||        :::      |||   :::   ||||||    ||||||
Db   4314 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4373

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
          |||               |||::::::|||   |||     ::::::|||||||||||     :::
Db   4374 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4433

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
             :::   |||   ||||||    |||   |||:::   :::       ||||||
Db   4434 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4493

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
          |||   ||||||||||||:::   |||      |||||||||||:::         ||||||
Db   4494 ATTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4553

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
          ||||||                :::                  ||||||   ||||||   |||
Db   4554 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4613

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
          |||:::   ||||||:::|||   |||:::||||||:::         :::         |||
Db   4614 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4673

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
                   ||||||:::||||||   :::   |||   :::   :::|||:::|||   |||
Db   4674 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4733

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
          :::      |||      :::         |||   |||   |||||||||      :::
Db   4734 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4784

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
                ||||||||     |||||||||||||:::||||||:::|||:::|||
Db   4785 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4838

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
          |||:::||||||||||||||||||||||||   ||||||   |||||||||     |||
Db   4839 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4898

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
                   :::   ||||||||         :::   ||||||::::::
Db   4899 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4958

Qy   1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
          :::         |||                    |||   :::|||||||||     ::
Db   4959 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5018

Qy   1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
           :    ||||||:::                       |||       :::|||
Db   5019 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5078

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
            :::         :::|||:::|||            |||:::    |||    |||:::
Db   5079 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5136
```

RESULT 13
AF001504

```
LOCUS       AF001504                6297 bp    mRNA    linear   PRI 06-JUN-1997
DEFINITION  Homo sapiens myasthenia gravis autoantigen gravin mRNA, partial
            cds.
ACCESSION   AF001504
VERSION     AF001504.1  GI:2160783
KEYWORDS    .
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 6297)
  AUTHORS   Bowditch,R.D. and Ginsberg,M.H.
  TITLE     Sequence of gravin cDNA isolated from a human umbilical vein
            endothelial cell library
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 6297)
  AUTHORS   Bowditch,R.D. and Ginsberg,M.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-APR-1997) Biochemistry, University of Oklahoma Health
            Sciences Center, PO Box 26901, BMSB 853, Oklahoma City, OK 73190,
            USA
FEATURES             Location/Qualifiers
     source          1..6297
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /cell_type="endothelial cell"
                     /tissue_type="umbilical vein"
     CDS             <1..5226
                     /note="kinase scaffold protein"
                     /codon_start=1
                     /product="myasthenia gravis autoantigen gravin"
                     /protein_id="AAB58938.1"
                     /db_xref="GI:2160784"
                     /translation="DPAIAASDPATKLLQKNGQLSTINGVAEQDELSLQEGDLNGQKG
                     ALNGQGALNSQEEEEVIVTEVGQRDSEDVSERDSDKEMATKSAVVHDITDDGQEETPE
                     IIEQIPSSESNLEELTQPTESQANDIGFKKVFKFVGFKFTVKKDKTEKPDTVQLLTVK
                     KDEGEGAAGAGDHQDPSLGAGEAASKESEPKQSTEKPEETLKREQSHAEISPPAESGQ
                     AVEECKEEGEEKQEKEPSKSAESPTSPVTSETGSTFKKFFTQGWAGWRKKTSFRKPKE
                     DEVEASEKKKEQEPEKVDTEEDGKAEVASEKLTASEQAHPQEPAESAHEPRLSAEYEK
                     VELPSEEQVSGSQGPSEEKPAPLATEVFDEKIEVHQEEVVAEVHVSTVEERTEEQKTE
                     VEETAGSVPAEELVEMDAEPQEAEPAKELVKLKETCVSGEDPTQGADLSPDEKVLSKP
                     PEGVVSEVEMLSSQERMKVQGSPLKKLFTSTGLKKLSGKKQKGKRGGGDEESGEHTQV
                     PADSPDSQEEQKGESSASSPEEPEEITCLEKGLAEVQQDGEAEEGATSDGEKKREGVT
                     PWASFKKMVTPKKRVRRPSESDKEDELDKVKSATLSSTESTASEMQEEMKGSVEEPKP
                     EEPKRKVDTSVSWEALICVGSSKKRARRGSSSDEEGGPKAMGGDHQKADEAGKDKETG
                     TDGILAGSQEHDPGQGSSSPEQAGSPTEGEGVSTWESFKRLVTPRKKSKSKLEEKSED
```

```
                     SIAGSGVEHSTPDTEPGKEESWVSIKKFIPGRRKKRPDGKQEQAPVEDAGPTGANEDD
                     SDVPAVVPLSEYDAVEREKMEAQQAQKSAEQPEQKAATEVSKELSESQVHMMAAAVAD
                     GTRAATIIEERSPSWISASVTEPLEQVEAEAALLTEEVLEREVIAEEEPPTVTEPLPE
                     NREARGDTVVSEAELTPEAVTAAETAGPLGAEEGTEASAAEETTEMVSAVSQLTDSPD
                     TTEEATPVQEVEGGVPDIEEQERRTQEVLQAVAEKVKEESQLPGTGGPEDVLQPVQRA
                     EAERPEEQAEASGLKKETDVVLKVDAQEAKTEPFTQGKVVGQTTPESFEKAPQVTESI
                     ESSELVTTCQAETLAGVKSQEMVMEQAIPPDSVETPTDSETDGSTPVADFDAPGTTQK
                     DEIVEIHEENEVASGTQSGGTEAEAVPAQKERPPAPSSFVFQEETKEQSKMEDTLEHT
                     DKEVSVETVSILSKTEGTQEADQYADEKTKDVPFFEGLEGSIDTGITVSREKVTEVAL
                     KGEGTEEAECKKDDALELQSHAKSPPSPVEREMVVQVEREKTEAEPTHVNEEKLEHET
                     AVTVSEEVSKQLLQTVNVPIIDGAKEVSSLEGSPPPCLGQEEAVCTKIQVQSSEASFT
                     LTAAAEEEKVLGETANILETGETLEPAGAHLVLEEKSSEKNEDFAAHPGEDAVPTGPD
                     CQAKSTPVIVSATTKKGLSSDLEGEKTTSLKWKSDEVDEQVACQEVKVSVAIEEDLEP
                     ENGILELETKSSKLVQNIIQTAVDQFVRTEETATEMLTSELQTQAHVIKADSQDAGQE
                     TEKEGEEPQASAQDETPITSAKEESESTAVGQAHSDISKDMSEASEKTMTVEVEGSTV
                     NDQQLEEVVLPSEEEGGGAGTKSVPEDDGHALLAERIEKSLVEPKEDEKGDDVDDPEN
                     QNSALADTDASGGLTKESPDTNGPKQKEKEDAQEVELQEGKVHSESDKAITPQAQEEL
                     QKQERESAKSELTES"
BASE COUNT     2051 a   1310 c   1708 g   1228 t
ORIGIN

Alignment Scores:
Pred. No.:              5.17e-143      Length:         6297
Score:                  4231.50        Matches:        943
Percent Similarity:     69.07%         Conservative:   187
Best Local Similarity:  57.64%         Mismatches:     398
Query Match:            52.42%         Indels:         108
DB:                     9              Gaps:           31

US-09-902-432-4 (1-1596) x AF001504 (1-6297)

Qy     41 AspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGlyGlnLeu 58
Db      1 GACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGTCAGCTG 60

Qy     59 SerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGlu 78
Db     61 TCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGACCTAAAT 120

Qy     79 GlyGln------------------------------------------GluGluGluVal 84
Db    121 GGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAAGAAGTC 180

Qy     85 ValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArgValGlu 104
Db    181 ATTGTCACAGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCCGATAAA 240

Qy    105 GluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGluGluThr 124
Db    241 GAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAGGAGACA 300

Qy    125 SerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnPro 144
Db    301 CCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCC 360

Qy    145 AlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGlyPheLys 164
Db    361 ACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAA 420

Qy    165 PheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThrValLys 184
Db    421 TTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAG 480

Qy    185 LysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSer 204
Db    481 AAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGACCCCAGC 531

Qy    205 ValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSerThrGlu 224
Db    532 CTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAG 585

Qy    225 LysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGlnAlaGlu 244
Db    586 AAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCAGCCGAA 645

Qy    245 SerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGlu 264
Db    646 TCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAA 702

Qy    265 ProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSerSerPhe 284
Db    703 CCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTC 762

Qy    285 LysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSer 304
Db    763 AAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCG 822

Qy    305 LysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAsp 324
Db    823 AAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGAC 882

Qy    325 GluGluGluLysGluLysThrGluProAlaSerGluGlu--------------------- 337
Db    883 ACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCC 942

Qy    338 ------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLys 355
Db    943 CACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAA 1002

Qy    356 ValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCys 375
Db   1003 GTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCT 1062

Qy    376 AlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---GluValVal 394
Db   1063 GCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTG 1122

Qy    395 AlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGlu 414
Db   1123 GCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------------- 1167

Qy    415 AlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAla 434
Db   1168 ---------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTT 1218

Qy    435 GluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGlu 454
Db   1219 GAAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAA 1278

Qy    455 MetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGluLysThr 474
Db   1279 ACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTG 1338

Qy    475 LeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGlnGluArg 494
Db   1339 CTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGA 1398

Qy    495 IleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeu 514
Db   1399 ATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAAAAGCTT 1458

Qy    515 SerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyr 534
Db   1459 TCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGGGAGCAC 1515

Qy    535 GlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAla 554
Db   1516 ACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCC 1575

Qy    555 SerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaPro 574
Db   1576 TCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAG 1635

Qy    575 GlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIle 594
Db   1636 CAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTC 1695
```

```
Qy    595 ThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArgProSer 614
Db   1696 ACTCCCTGGGCTTCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGGCCTTCG 1755

Qy    615 GluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAsp 634
Db   1756 GAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAG 1815

Qy    635 SerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGlu 654
Db   1816 AGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAGCCGGAA 1875

Qy    655 GluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySer 674
Db   1876 GAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCA 1935

Qy    675 SerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeu 694
Db   1936 TCCAAGAAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATG 1995

Qy    695 GlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAla 714
Db   1996 GGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGG 2055

Qy    715 ValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAla 734
Db   2056 ATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCT 2115

Qy    735 GlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThr 754
Db   2116 GGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACG 2175

Qy    755 ProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp------------Ser 770
Db   2176 CCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCT 2235

Qy    771 SerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIle 790
Db   2236 GGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATC 2295

Qy    791 LysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThr 810
Db   2296 AAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCT 2355

Qy    811 ValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValVal 830
Db   2356 GTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTC 2415

Qy    831 ProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------AlaGlnGly 847
Db   2416 CCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAA 2475

Qy    848 AsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThr 867
Db   2476 AGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGT 2535

Qy    868 LeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGlu 887
Db   2536 CAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAA 2595

Qy    888 GluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGly 907
Db   2596 GAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCT 2655

Qy    908 GluAlaMetProProValGluGluValThrGluLysAspIleIleAla---GluGluThr 926
Db   2656 GAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCC 2715

Qy    927 ProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSer 946
Db   2716 CCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGT 2775

Qy    947 GluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThr 966
Db   2776 GAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTGGGTGCC 2835

Qy    967 GluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSer 986
Db   2836 GAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCAGTCTCC 2895

Qy    987 GlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGluSer 1006
Db   2896 CAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTGGAAGGT 2955

Qy   1007 GlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAla 1026
Db   2956 GGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCAGTGGCA 3015

Qy   1027 AspLysValLysGluGluSerGlnValProAlaThr---------------------Gln 1039
Db   3016 GAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTGCTTCAG 3075

Qy   1040 ThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSer 1059
Db   3076 CCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAGCAGGCT 3117

Qy   1060 GluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAla 1079
Db   3118 GAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAGGAGGCA 3177

Qy   1080 GlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGlu 1099
Db   3178 AAAACTGAGCCTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGCTTTGAA 3237

Qy   1100 ---ValProGluValThrAlaAspValAspHis---------ValAlaThrCysGlnVal 1115
Db   3238 AAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGTCAAGCC 3297

Qy   1116 ---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAlaProGlu 1129
Db   3298 GAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCCCCTGAC 3357

Qy   1130 SerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAspSerAsp 1149
Db   3358 TCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGACTTTGAC 3417

Qy   1150 ThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAlaThrAla 1169
Db   3418 GCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAGGTCGCA 3477

Qy   1170 AlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluPro 1189
Db   3478 TCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAGAGGCCT 3537

Qy   1190 SerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly---Arg 1208
Db   3538 CCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAGATGGAA 3594

Qy   1209 AspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLys 1228
Db   3595 GACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTGTCAAAG 3654

Qy   1229 ThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys------ 1246
Db   3655 ACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAAGACGTA 3708

Qy   1247 ---------GluGluGlnGluValPheValHisSerGly------ProAsnSerGlnLys 1261
Db   3709 CCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGGGAAAAG 3768

Qy   1262 AlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGlu 1281
Db   3769 GTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAGGATGAT 3828

Qy   1282 SerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGluMetGlu 1296
```

```
           :::  |||:::||||||                          |||       |||    ||||||
Db    3829 GCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAGATGGTA 3888

Qy    1297 ThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGluGly--- 1315
              ||||||:::||||||     |||     :::|||       |||:::||||||
Db    3889 GTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAGAAGCTT 3948

Qy    1316 GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeu 1335
           |||   |||||||||            |||            :::   :::   |||:::
Db    3949 GAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAGACAGTG 4008

Qy    1336 AspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer---ProSer 1354
           ::::::|||       :::      |||   :::   ||||||   ||||||   |||
Db    4009 AATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCTCCTCCC 4068

Qy    1355 LeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThr 1374
                          |||::::::|||   |||   ::::::||||||||||||   :::   :::
Db    4069 TGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAGGCATCA 4128

Qy    1375 ValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValValIleSer 1392
              |||   ||||||   |||   |||:::   :::       ||||||      |||
Db    4129 TTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAACATTTTA 4188

Qy    1393 GluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLysSerSer 1412
           ||||||||||||:::   |||      ||||||||||||:::      ||||||||||||
Db    4189 GAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAATCCTCT 4248

Qy    1413 AlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyProGlu 1432
                              :::               ||||||   ||||||   ||||||:::
Db    4249 GAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGGCCCGAC 4308

Qy    1433 SerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisPro 1452
              ||||||:::|||   |||:::||||||:::            :::      |||
Db    4309 TGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTAAGTTCC 4368

Qy    1453 AspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysPro 1472
           ||||||:::||||||      :::   |||      :::   :::|||:::|||   |||:::
Db    4369 GACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGATGAGCAG 4428

Qy    1473 AspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeuLysAla 1492
              |||      :::            |||   |||   |||||||||:::   |||:::
Db    4429 GTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTAGAGCCT 4482

Qy    1493 GluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnValIleGln 1512
           |||         |||||||||||||||:::||||||:::|||:::|||   |||:::||||||
Db    4483 GAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATCATCCAG 4542

Qy    1513 ThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThrHisAlaTyrAsp 1531
           ||||||||||||||||||   ||||||   |||||||||   |||            :::
Db    4543 ACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACGTCTGAG 4602

Qy    1532 SerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCysTrpThr 1551
              |||||||||            :::   ||||||::::::         :::
Db    4603 TTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAAACGGAG 4662

Qy    1552 Lys---------------MetLysAspAla-LysMetLysHisProValProGlnProAr 1566
           |||                |||  :::||||||||||    :::    ||||||::
Db    4663 AAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCAGCCAAA 4722

Qy    1566 gGlu---------------------AspLeuGlnValLeuThrValLeuGluAlaTrpAl 1579
           :                         |||      :::|||         :::
Db    4723 GAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGACATGAGT 4782

Qy    1579 aGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
             :::|||:::|||            |||:::   |||   |||:::
Db    4783 GAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 4828
```

```
RESULT 14
AL590413
LOCUS       AL590413               104939 bp    DNA     linear   PRI 06-DEC-2001
DEFINITION  Human DNA sequence from clone RP11-351K16 on chromosome 6, complete
            sequence.
ACCESSION   AL590413
VERSION     AL590413.18  GI:17426497
KEYWORDS    HTG.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 104939)
  AUTHORS   Martin,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (06-DEC-2001) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Dec 8, 2001 this sequence version replaced gi:17384482.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
            was generated from part of bacterial clone contigs of human
            chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
            Group.  Further information can be found at
            http://www.sanger.ac.uk/HGP/Chr6
            RP11-351K16 is from the library RPCI-11.2 constructed by the group
            of Pieter de Jong. For further details see
            http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6
            IMPORTANT: This sequence is not the entire insert of clone
            RP11-351K16 It may be shorter because we sequence overlapping
            sections only once, except for a short overlap.
            The true right end of clone RP11-351K16 is at 104939 in this
            sequence. The true left end of clone RP11-108N8 is at 59858 in this
            sequence. The true right end of clone RP3-403M6 is at 2000 in this
            sequence.
FEATURES             Location/Qualifiers
     source          1..104939
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /clone="RP11-351K16"
                     /clone_lib="RPCI-11.2"
     misc_feature    51026..51045
                     /note="Sequence from AC023201 sequenced by WIBR."
     misc_feature    51799..51952
                     /note="Single clone region. Sequence from reads from a
                     short insert library derived from a single pUC clone.
                     Restriction digest data confirm the assembly."
     misc_feature    52345..52447
                     /note="Sequence from reads from a short insert library
                     derived from a single pUC clone.  Restriction digest data
                     confirm the assembly."
     misc_feature    86009..86146
                     /note="Sequence from overlapping clone RP11-108N8
                     (AL590543).  Assembly confirmed by restriction digest."
     misc_feature    87770..87807
                     /note="Sequence from uni-directional dGTP big dye
                     terminator reads only."
BASE COUNT    30065 a  22273 c  22882 g  29719 t
ORIGIN

Alignment Scores:
```

```
Pred. No.:                  4.23e-137     Length:         104939
Score:                      4097.00       Matches:        917
Percent Similarity:         68.28%        Conservative:   181
Best Local Similarity:      57.03%        Mismatches:     391
Query Match:                50.75%        Indels:         119
DB:                         9             Gaps:           32

US-09-902-432-4 (1-1596) x AL590413 (1-104939)

Qy      77 GlnGluGlyGlnGluGluGluValValAspGluAsp------------------------ 88
           |||   ||||||||   |||    :::   |||
Db    9313 CAGCCGGGGCAAGAGAGTGAGGCCCTGTCGGAGAAAAACAATTTCTTGTATTGTAATCA 9372

Qy      89 ----------------------------------------ValGlyGlnArgGluSer 94
                                                   ||||||||||||:::|||
Db    9373 CCTTTTCTCTTCTCCCACCCCCCGCCCCTTTTTGTTAATAGTTGGACAGAGAGACTCT 9432

Qy      95 GluAspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGlu 114
           ||||||||    ::::::|||      :::||||||||        |||       |||
Db    9433 GAAGATGTGAGCAAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCAC 9492

Qy     115 AspIleThrLysAspGlyGlnGluGluThrSerGluIleIleGluGlnIleProAlaSer 134
           ||||||||  ||||||||||||||||||    |||||||||||||||||||||:::|||
Db    9493 GACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCTTCTTCA 9552

Qy     135 GluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPhe 154
           |||:::|||:::||||||:::      ||||||   ||||||||||||||||:::||||||
Db    9553 GAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATTGGATTT 9612

Qy     155 LysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLys 174
           ||||||||||||||||||||||||||||||||||||||||||||||||||     ||||||
Db    9613 AAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAG 9672

Qy     175 SerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSer 194
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9673 CCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCA--------- 9723

Qy     195 ValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSer 214
           ||||||||||||||||||:::::||||||:::                ||||||:::||||||
Db    9724 GCAGGGGCTGGCGACCACAAGGACCCCAGCCTT------GGGGCTGGAGAAGCAGCATCC 9777

Qy     215 LysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGln 234
           ||||||||||||   |||||||||||||||||   |||   |||||||||:::||||||
Db    9778 AAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAA 9837

Qy     235 SerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLys 254
           |||       ||||||         |||||||||   ||||||      ||||||   |||
Db    9838 AGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAA 9894

Qy     255 AspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerProSerSer 274
           :::||||||||||||||||||||||||||||:::||||||   |||||||||:::|||
Db    9895 GAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGT 9954

Qy     275 ProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGly 294
           ||||||   |||||||||   |||:::||||||||||||||||||   ||||||||||||
Db    9955 CCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGC 10014

Qy     295 TrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLys 314
           |||||||||||||||||||||:::||||||||||||||||:::::::|||   :::||||||
Db   10015 TGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAG 10074

Qy     315 ArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGluLysThrGluProAla 334
           :::|||||||||||||   ||||||||||||   ||||||      |||   |||   |||
Db   10075 AAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCC 10134

Qy     335 SerGluGlu---------------------------GlnGluProAlaGluAspThrAsp 345
           ||||||:::                           ||||||||||||||||||
Db   10135 TCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCAC 10194

Qy     346 GlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGly 365
           :::      |||||||||||:::|||||||||||||||||||      |||:::||||||
Db   10195 GAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGT 10254

Qy     366 AspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGlu 385
               :::         |||||||||||   |||||||||||||||||||||||||||||
Db   10255 GGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAG 10314

Qy     386 LysMetGluAlaHisGln---GluValValAlaGluValHisValSerThrValGluLys 404
           |||:::|||   ||||||   ||||||||||||||||||||||||||||||||||||:::
Db   10315 AAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAG 10374

Qy     405 ThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThr 424
              |||||||||                                 ||||||    |||
Db   10375 AGAACCGAAGAGCAG-----------------------AAAACGGAGGTGGAAGAAACA 10410

Qy     425 GlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGlu 444
                 |||:::|||   |||:::|||   |||         ||||||||||||||
Db   10411 GCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAAGCTGAA 10470

Qy     445 ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGln 464
           ||||||:::||||||:::|||      :::|||   ||||||||||||   |||   ||||||
Db   10471 CCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAG 10530

Qy     465 LeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleValSer 484
                 ||||||||||||:::||||||   |||   |||   |||||||||:::||||||
Db   10531 GGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGT 10590

Qy     485 GluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLys 504
           ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||||
Db   10591 GAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAG 10650

Qy     505 LeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGly 524
           ||||||:::|||:::|||||||||||||||||||||||||||||||||||||||||||
Db   10651 CTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA--- 10707

Qy     525 GlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSer 544
           ||||||||||||||||||   ||||||:::         :::      :::||||||:::|||
Db   10708 GGAGGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGC 10767

Qy     545 AlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThr 564
              :::||||||||||||||||||||||||||||||||||||||||||||||   |||
Db   10768 CAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACG 10827

Qy     565 CysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThr 584
           ||||||||||||||         |||      ||||||||||||||||||      |||
Db   10828 TGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACT 10887

Qy     585 SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetVal 604
           |||||||||||||||||||||||:::||||||||||||||||||||||||||||||||
Db   10888 TCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTG 10947

Qy     605 ThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLys 624
           ||||||||||||||||||||||||||||||||||||||||:::||||||:::|||
Db   10948 ACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAG 11007

Qy     625 ValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluVal 644
           ||||||||||||||||||||||||:::||||||      ||||||||||||:::|||:::
Db   11008 GTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATG 11067

Qy     645 LysThrValGlyGluGluGlnLysProGluGluProLysArgArgValAspThrSerVal 664
           |||         ||||||      |||||||||||||||:::||||||||||||||||
Db   11068 AAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTA 11127

Qy     665 SerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSer 684
           ||||||||||||||||||||||||||||||||||||||||||||||:::      ||||||
Db   11128 TCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGGTCCTCT 11187

Qy     685 SerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAla 704
           ||||||:::||||||||||||:::      :::||||||||      :::|||:::||||||
Db   11188 TCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCC 11247

Qy     705 SerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGln 724
              ||||||||||||   ||||||||||   :::   |||   :::||||||   |||
Db   11248 GGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACATGATCCA 11307

Qy     725 AlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSer 744
```

```
Db  11308 GGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCC 11367

Qy    745 ThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGlu 764
Db  11368 ACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAA 11427

Qy    765 GluLysAlaGluAsp------------SerSerValGluGlnLeuSerThrGluIleGlu 780
Db  11428 GAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGACACTGAA 11487

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
Db  11488 CCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAA 11547

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
Db  11548 AGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAAC 11607

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
Db  11608 GAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGG 11667

Qy    841 GluLysMetGlu---------AlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAla 857
Db  11668 GAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCC 11727

Qy    858 ValTyrValSerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIle 877
Db  11728 ACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCT 11787

Qy    878 AspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSer 897
Db  11788 GACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCA 11847

Qy    898 ValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThr 917
Db  11848 GTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTG 11907

Qy    918 GluLysAspIleIleAla---GluGluThrProValLeuThrGlnThrLeuProGluGly 936
Db  11908 GAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCAGAGAAC 11967

Qy    937 LysAspAlaHisAspAspMetValThrSerGluValAspPheThrSerGluAlaValThr 956
Db  11968 AGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACA 12027

Qy    957 AlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGlu 976
Db  12028 GCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAA 12087

Qy    977 GluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGlu 996
Db  12088 GAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACCACAGAG 12147

Qy    997 GluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluArg 1016
Db  12148 GAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGG 12207

Qy   1017 GlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValPro 1036
Db  12208 CGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCT 12267

Qy   1037 AlaThr--------------------GlnThrValGlnArgThrGlySerLysAlaLeu 1049
Db  12268 GGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCA--------------- 12312

Qy   1050 GluLysValGluGluValGluGluAspSerGluValLeuAlaSerGluLysGluLysAsp 1069
Db  12313 ---GAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGAT 12369

Qy   1070 ValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGlu 1089
Db  12370 GTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTG 12429

Qy   1090 ThrGlyGlnAlaThrProGluSerLeuGlu---ValProGluValThrAlaAspValAsp 1108
Db  12430 GTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAG 12489

Qy   1109 His---------ValAlaThrCysGlnVal---------------IleLysLeuGlnGln 1120
Db  12490 TCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAG 12549

Qy   1121 Leu---MetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThr 1139
Db  12550 ATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACT 12609

Qy   1140 AsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThr 1159
Db  12610 GATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATT 12669

Qy   1160 IleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGlu 1179
Db  12670 GTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCA 12729

Qy   1180 GluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGln 1199
Db  12730 GAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAG 12786

Qy   1200 GluGluHisGlyGluGluProGly---ArgAspValLeuGluProThrGlnGlnGluLeu 1218
Db  12787 GAAGAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTG 12846

Qy   1219 ThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAsp 1238
Db  12847 TCAGTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACTCAA------GAGGCTGAC 12900

Qy   1239 TrpLeuAspGlyGluLysValLys---------------GluGluGlnGluValPheVal 1253
Db  12901 CAGTATGCTGATGAGAAAACCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATA 12960

Qy   1254 HisSerGly------ProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGluVal 1271
Db  12961 GACACAGGCATAACAGTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGG 13020

Qy   1272 MetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSer------------ 1287
Db  13021 ACAGAAGAAGCTGAATGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCT 13080

Qy   1288 ---LeuSerLeuGluGlyGluMetGluThrAspValGluLysGluLysArgGluThr 1306
Db  13081 CCTCCATCCCCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCA 13140

Qy   1307 LysProGluGlnValSerGluGluGly---GluGlnGluThrAlaAlaProGluHisGlu 1325
Db  13141 GAGCCAACCCATGTGAATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAA 13200

Qy   1326 GlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAla 1345
Db  13201 GAGGTCAGTAAGCAGCTCCTCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAA 13260

Qy   1346 LeuGlySerLeuGlyGlySer---ProSerLeuProAspGlnAspLysAlaGlyCysIle 1364
Db  13261 GTCAGCAGTTTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACC 13320

Qy   1365 GluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLys 1384
Db  13321 AAAATTCAAGTTCAGAGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAA 13380

Qy   1385 ValIle------GluThrValValIleSerGluThrGlyGluSerProGluCysValGly 1402
Db  13381 AAGGTCTTAGGAGAAACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGT 13440

Qy   1403 AlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHis 1422
Db  13441 GCACATTTAGTTCTGGAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCA 13500
```

```
Qy    1423 AlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleVal 1442
              ||||||   ||||||   ||||||:::   ||||||:::|||   |||:::||||||
Db   13501 GGGGAAGATGCTGTGCCCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTA 13560

Qy    1443 ThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGln 1462
           :::         :::        |||      ||||||:::||||||   :::   |||
Db   13561 TCTGCTACTACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTG 13620

Qy    1463 ArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGlu 1482
           :::   :::|||:::|||   |||:::         |||      :::         |||
Db   13621 AAGTGGAAGTCAGATGAAGTCGATGAGCAGGTTGCTTGCCAGGAGGTC------AAAGTG 13674

Qy    1483 SerThrAlaIleGluLysValLeuLysAlaGluProGlu------IleLeuGluLeuGlu 1500
           |||   |||||||||   :::         |||||||||      |||||||||||||||
Db   13675 AGTGTAGCAATTGAGGATTTA---------GAGCCTGAAAATGGGATTTTGGAACTTGAG 13725

Qy    1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
           :::||||||:::|||:::|||   |||:::||||||||||||||||||||||   |||
Db   13726 ACCAAAAGCAGTAAACTTGTCCAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGT 13785

Qy    1521 Thr---GluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCys 1539
           |||   |||||||||   |||               :::   |||||||||
Db   13786 ACAGAAGAAACAGCCACCGAAATGTTGACGTCTGAGTTACAGACACAAGCTCACGTGATA 13845

Qy    1540 ArgLeuAspSerArgGluProAsnArgCysTrpThrLys---------------MetLys 1554
           :::   ||||||::::::         :::         |||
Db   13846 AAAGCTGACAGCCAGGACGCTGGACAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCC 13905

Qy    1555 AspAla-LysMetLysHisProValProGlnProArgGlu-------------------- 1567
              ||| :::|||||||||   :::   ||||||:::
Db   13906 TCTGCACAGGATGAAACACCAATTACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTG 13965

Qy    1568 -AspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLeuProAr 1587
             |||      :::|||         :::         :::|||:::|||
Db   13966 GGACAAGCACATTCTGATATTTCCAAAGACATGAGTGAAGCCTCAGAAAAGACCATGACT 14025

Qy    1587 gLeuGlnLeuLysAlaProVal 1594
             |||:::   |||   |||:::
Db   14026 GTTGAGGTAGAAGGTTCCACTG 14047
```

RESULT 15
AC023201

```
LOCUS       AC023201               171158 bp    DNA     linear   HTG 24-AUG-2002
DEFINITION  Homo sapiens chromosome 6 clone RP11-193D24 map 6, WORKING DRAFT
            SEQUENCE, 26 unordered pieces.
ACCESSION   AC023201
VERSION     AC023201.2  GI:7139764
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 171158)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 6, clone RP11-193D24
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 171158)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
            McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
            Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
            Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
            Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
            Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (09-FEB-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 171158)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 1, 2000 this sequence version replaced gi:6957769.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------------- Genome Center
                Center: Whitehead Institute/ MIT Center for Genome Research
                Center code: WIBR
                Web site: http://www-seq.wi.mit.edu
                Contact: sequence_submissions@genome.wi.mit.edu
            -------------- Project Information
                Center project name: L6150
                Center clone name: 193_D_24
            -------------- Summary Statistics
                Sequencing vector: M13; M77815; 100% of reads
                Chemistry: Dye-terminator Big Dye; 100% of reads
                Assembly program: Phrap; version 0.960731
                Consensus quality: 159689 bases at least Q40
                Consensus quality: 164965 bases at least Q30
                Consensus quality: 167080 bases at least Q20
                Insert size: 176000; agarose-fp
                Insert size: 168658; sum-of-contigs
                Quality coverage: 4.8 in Q20 bases; agarose-fp
                Quality coverage: 5.0 in Q20 bases; sum-of-contigs
            --------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 26 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1      1187: contig of 1187 bp in length
            *     1188 1287: gap of       100 bp
            *     1288      3150: contig of 1863 bp in length
            *     3151 3250: gap of       100 bp
            *     3251      4999: contig of 1749 bp in length
            *     5000 5099: gap of       100 bp
            *     5100      6170: contig of 1071 bp in length
            *     6171 6270: gap of       100 bp
            *     6271      7887: contig of 1617 bp in length
            *     7888 7987: gap of       100 bp
            *     7988      9424: contig of 1437 bp in length
```

```
        *     9425 9524: gap of      100 bp
        *     9525    11769: contig of 2245 bp in length
        *    11770 11869: gap of      100 bp
        *    11870    13231: contig of 1362 bp in length
        *    13232 13331: gap of      100 bp
        *    13332    14939: contig of 1608 bp in length
        *    14940 15039: gap of      100 bp
        *    15040    17089: contig of 2050 bp in length
        *    17090 17189: gap of      100 bp
        *    17190    19635: contig of 2446 bp in length
        *    19636 19735: gap of      100 bp
        *    19736    23022: contig of 3287 bp in length
        *    23023 23122: gap of      100 bp
        *    23123    27687: contig of 4565 bp in length
        *    27688 27787: gap of      100 bp
        *    27788    32731: contig of 4944 bp in length
        *    32732 32831: gap of      100 bp
        *    32832    37404: contig of 4573 bp in length
        *    37405 37504: gap of      100 bp
        *    37505    42846: contig of 5342 bp in length
        *    42847 42946: gap of      100 bp
        *    42947    50280: contig of 7334 bp in length
        *    50281 50380: gap of      100 bp
        *    50381    57865: contig of 7485 bp in length
        *    57866 57965: gap of      100 bp
        *    57966    66245: contig of 8280 bp in length
        *    66246 66345: gap of      100 bp
        *    66346    74797: contig of 8452 bp in length
        *    74798 74897: gap of      100 bp
        *    74898    84807: contig of 9910 bp in length
        *    84808 84907: gap of      100 bp
        *    84908    96491: contig of 11584 bp in length
        *    96492 96591: gap of      100 bp
        *    96592   111786: contig of 15195 bp in length
        *   111787 111886: gap of      100 bp
        *   111887   128020: contig of 16134 bp in length
        *   128021 128120: gap of      100 bp
        *   128121   145273: contig of 17153 bp in length
        *   145274 145373: gap of      100 bp
        *   145374   171158: contig of 25785 bp in length.
FEATURES             Location/Qualifiers
     source          1. .171158
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="6"
                     /map="6"
                     /clone="RP11-193D24"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1. .1187
                     /note="assembly_fragment"
     misc_feature    1288. .3150
                     /note="assembly_fragment"
     misc_feature    3251. .4999
                     /note="assembly_fragment"
     misc_feature    5100. .6170
                     /note="assembly_fragment"
     misc_feature    6271. .7887
                     /note="assembly_fragment"
     misc_feature    7988. .9424
                     /note="assembly_fragment"
     misc_feature    9525. .11769
                     /note="assembly_fragment"
     misc_feature    11870. .13231
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
     misc_feature    13332. .14939
                     /note="assembly_fragment"
     misc_feature    15040. .17089
                     /note="assembly_fragment"
     misc_feature    17190. .19635
                     /note="assembly_fragment"
     misc_feature    19736. .23022
                     /note="assembly_fragment"
     misc_feature    23123. .27687
                     /note="assembly_fragment"
     misc_feature    27788. .32731
                     /note="assembly_fragment"
     misc_feature    32832. .37404
                     /note="assembly_fragment"
     misc_feature    37505. .42846
                     /note="assembly_fragment"
     misc_feature    42947. .50280
                     /note="assembly_fragment"
     misc_feature    50381. .57865
                     /note="assembly_fragment"
     misc_feature    57966. .66245
                     /note="assembly_fragment"
     misc_feature    66346. .74797
                     /note="assembly_fragment"
     misc_feature    74898. .84807
                     /note="assembly_fragment"
     misc_feature    84908. .96491
                     /note="assembly_fragment"
     misc_feature    96592. .111786
                     /note="assembly_fragment"
     misc_feature    111887. .128020

Alignment Scores:
Pred. No.:               6.59e-137      Length:         171158
Score:                   4097.00        Matches:        917
Percent Similarity:      68.28%         Conservative:   181
Best Local Similarity:   57.03%         Mismatches:     391
Query Match:             50.75%         Indels:         119
DB:                      2              Gaps:           32

US-09-902-432-4 (1-1596) x AC023201 (1-171158)

Qy       77 GlnGluGlyGlnGluGluGluValValAspGluAsp------------------------ 88
             |||  ||||||||  |||   :::  |||
Db    87301 CAGCCGGGGCAAGAGAGTGAGGCCCTGTCGGAGAAAAACAATTTTCTTGTATTGTAATCA 87360

Qy       89 ------------------------------------------ValGlyGlnArgGluSer 94
                                                      |||||||||||:::|||
Db    87361 CCTTTTCTCTTCTCCCCACCCCCCCGCCCCTTTTTGTTAATAGTTGGACAGAGAGACTCT 87420

Qy       95 GluAspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGlu 114
            ||||||||    ::::::|||      :::|||||||||      |||      |||
Db    87421 GAAGATGTGAGCAAAAGAGACTCCGATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCAC 87480

Qy      115 AspIleThrLysAspGlyGlnGluGluThrSerGluIleIleGluGlnIleProAlaSer 134
            ||||||||   ||||||||||||||||||   |||||||||||||||||||:::|||
Db    87481 GACATCACAGATGATGGGCAGGAGGAGACACCCGAAATAATCGAACAGATTCCTTCTTCA 87540

Qy      135 GluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPhe 154
            |||:::|||:::||||||:::     ||||||   ||||||||||||||||:::||||||
Db    87541 GAAAGCAATTTAGAAGAGCTAACACAACCCACTGAGTCCCAGGCTAATGATATTGGATTT 87600

Qy      155 LysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLys 174
            |||||||||||||||||||||||||||||||||||||||||||||||||||   ||||||
Db    87601 AAGAAGGTGTTTAAGTTTGTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAG 87660

Qy      175 SerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSer 194
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    87661 CCTGACACTGTCCAGCTACTCACTGTGAAGAAAGATGAAGGGGAGGGAGCA--------- 87711

Qy      195 ValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSer 214
               ||||||||||||||:::::::|||||||:::             ||||||:::||||||
Db    87712 GCAGGGGCTGGCGACCACAAGGACCCCAGCCTT------GGGGCTGGAGAAGCAGCATCC 87765

Qy      215 LysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGln 234
            |||||||||||   ||||||||||||||||||   |||   |||||||||:::||||||
Db    87766 AAAGAAAGCGAACCCAAACAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAA 87825

Qy      235 SerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLys 254
            |||      ||||||         |||||||||   ||||||      ||||||   |||
```

```
Db  87826 AGCCACGCAGAAATTTCTCCCCCAGCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAA 87882

Qy    255 AspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerProSerSer 274
Db  87883 GAGGAAGGAGAAGAGAAACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGT 87942

Qy    275 ProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGly 294
Db  87943 CCCGTGACCAGTGAAACAGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGC 88002

Qy    295 TrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLys 314
Db  88003 TGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAG 88062

Qy    315 ArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGluLysThrGluProAla 334
Db  88063 AAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCC 88122

Qy    335 SerGluGlu--------------------------GlnGluProAlaGluAspThrAsp 345
Db  88123 TCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCAC 88182

Qy    346 GlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGly 365
Db  88183 GAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGT 88242

Qy    366 AspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGlu 385
Db  88243 GGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAG 88302

Qy    386 LysMetGluAlaHisGln---GluValValAlaGluValHisValSerThrValGluLys 404
Db  88303 AAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAG 88362

Qy    405 ThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThr 424
Db  88363 AGAACCGAAGAGCAG-----------------------AAAACGGAGGTGGAAGAAACA 88398

Qy    425 GlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGlu 444
Db  88399 GCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAAGCTGAA 88458

Qy    445 ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGln 464
Db  88459 CCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAG 88518

Qy    465 LeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleValSer 484
Db  88519 GGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGT 88578

Qy    485 GluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLys 504
Db  88579 GAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAG 88638

Qy    505 LeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGly 524
Db  88639 CTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA--- 88695

Qy    525 GlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSer 544
Db  88696 GGAGGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGC 88755

Qy    545 AlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThr 564
Db  88756 CAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACG 88815

Qy    565 CysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThr 584
Db  88816 TGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACT 88875

Qy    585 SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetVal 604
Db  88876 TCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTG 88935

Qy    605 ThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLys 624
Db  88936 ACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAG 88995

Qy    625 ValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluVal 644
Db  88996 GTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATG 89055

Qy    645 LysThrValGlyGluGluGlnLysProGluGluProLysArgArgValAspThrSerVal 664
Db  89056 AAAGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTA 89115

Qy    665 SerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSer 684
Db  89116 TCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGTCCTCT 89175

Qy    685 SerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAla 704
Db  89176 TCTGATGAGGAAGGGGGACCAAAAGCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCC 89235

Qy    705 SerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGln 724
Db  89236 GGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACATGATCCA 89295

Qy    725 AlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSer 744
Db  89296 GGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCC 89355

Qy    745 ThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGlu 764
Db  89356 ACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAA 89415

Qy    765 GluLysAlaGluAsp------------SerSerValGluGlnLeuSerThrGluIleGlu 780
Db  89416 GAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGACACTGAA 89475

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
Db  89476 CCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAA 89535

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
Db  89536 AGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAAC 89595

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
Db  89596 GAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGG 89655

Qy    841 GluLysMetGlu---------AlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAla 857
Db  89656 GAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCC 89715

Qy    858 ValTyrValSerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIle 877
Db  89716 ACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCT 89775

Qy    878 AspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSer 897
Db  89776 GACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCA 89835

Qy    898 ValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThr 917
Db  89836 GTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTG 89895

Qy    918 GluLysAspIleIleAla---GluGluThrProValLeuThrGlnThrLeuProGluGly 936
Db  89896 GAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAACCTCTGCCAGAGAAC 89955

Qy    937 LysAspAlaHisAspAspMetValThrSerGluValAspPheThrSerGluAlaValThr 956
Db  89956 AGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACA 90015
```

```
Qy     957 AlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGlu 976
Db   90016 GCTGCAGAAACTGCAGGGCCATTGGGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAA 90075

Qy     977 GluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGlu 996
Db   90076 GAGACCACAGAAATGGTGTCAGCAGTCTCCCAGTTAACCGACTCCCCAGACACCACAGAG 90135

Qy     997 GluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArg 1016
Db   90136 GAGGCCACTCCGGTGCAGGAGGTGGAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGG 90195

Qy    1017 GlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValPro 1036
Db   90196 CGGACTCAAGAGGTCCTCCAGGCAGTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCT 90255

Qy    1037 AlaThr--------------------GlnThrValGlnArgThrGlySerLysAlaLeu 1049
Db   90256 GGCACCGGTGGGCCAGAAGATGTGCTTCAGCCTGTGCAGAGAGCA--------------- 90300

Qy    1050 GluLysValGluGluValGluGluAspSerGluValLeuAlaSerGluLysGluLysAsp 1069
Db   90301 ---GAGGCAGAAAGACCAGAAGAGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGAT 90357

Qy    1070 ValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGlu 1089
Db   90358 GTAGTGTTGAAAGTAGATGCTCAGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTG 90417

Qy    1090 ThrGlyGlnAlaThrProGluSerLeuGlu---ValProGluValThrAlaAspValAsp 1108
Db   90418 GTGGGGCAGACCACCCCAGAAAGCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAG 90477

Qy    1109 His---------ValAlaThrCysGlnVal---------------IleLysLeuGlnGln 1120
Db   90478 TCCAGTGAGCTTGTAACCACTTGTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAG 90537

Qy    1121 Leu---MetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThr 1139
Db   90538 ATGGTGATGGAACAGGCTATCCCCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACT 90597

Qy    1140 AsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThr 1159
Db   90598 GATGGAAGCACCCCCGTAGCCGACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATT 90657

Qy    1160 IleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGlu 1179
Db   90658 GTGGAAATCCATGAGGAGAATGAGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCA 90717

Qy    1180 GluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGln 1199
Db   90718 GAGGCAGTTCCTGCACAGAAAGAGAGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAG 90774

Qy    1200 GluGluHisGlyGluGluProGly---ArgAspValLeuGluProThrGlnGlnGluLeu 1218
Db   90775 GAAGAAACTAAAGAACAATCAAAGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTG 90834

Qy    1219 ThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAsp 1238
Db   90835 TCAGTGGAAACTGTATCCATTCTGTCAAAGACTGAGGGGACTCAA------GAGGCTGAC 90888

Qy    1239 TrpLeuAspGlyGluLysValLys---------------GluGluGlnGluValPheVal 1253
Db   90889 CAGTATGCTGATGAGAAAACCAAAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATA 90948

Qy    1254 HisSerGly------ProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGluVal 1271
Db   90949 GACACAGGCATAACAGTCAGTCGGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGG 91008

Qy    1272 MetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSer------------ 1287
Db   91009 ACAGAAGAAGCTGAATGTAAAAAGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCT 91068

Qy    1288 ---LeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThr 1306
Db   91069 CCTCCATCCCCCGTGGAGAGAGAGATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCA 91128

Qy    1307 LysProGluGlnValSerGluGluGly---GluGlnGluThrAlaAlaProGluHisGlu 1325
Db   91129 GAGCCAACCCATGTGAATGAAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAA 91188

Qy    1326 GlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAla 1345
Db   91189 GAGGTCAGTAAGCAGCTCCTCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAA 91248

Qy    1346 LeuGlySerLeuGlyGlySer---ProSerLeuProAspGlnAspLysAlaGlyCysIle 1364
Db   91249 GTCAGCAGTTTGGAAGGAAGCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACC 91308

Qy    1365 GluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLys 1384
Db   91309 AAAATTCAAGTTCAGAGCTCTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAA 91368

Qy    1385 ValIle------GluThrValValIleSerGluThrGlyGluSerProGluCysValGly 1402
Db   91369 AAGGTCTTAGGAGAAACTGCCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGT 91428

Qy    1403 AlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHis 1422
Db   91429 GCACATTTAGTTCTGGAAGAGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCA 91488

Qy    1423 AlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleVal 1442
Db   91489 GGGGAAGATGCTGTGCCCACAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTA 91548

Qy    1443 ThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGln 1462
Db   91549 TCTGCTACTACCAAGAAAGGCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTG 91608

Qy    1463 ArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGlu 1482
Db   91609 AAGTGGAAGTCAGATGAAGTCGATGAGCAGGTTGCTTGCCAGGAGGTC------AAAGTG 91662

Qy    1483 SerThrAlaIleGluLysValLeuLysAlaGluProGlu------IleLeuGluLeuGlu 1500
Db   91663 AGTGTAGCAATTGAGGATTTA---------GAGCCTGAAAATGGGATTTTGGAACTTGAG 91713

Qy    1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
Db   91714 ACCAAAAGCAGTAAACTTGTCCAAAACATCATCCAGACAGCCGTTGACCAGTTTGTACGT 91773

Qy    1521 Thr---GluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCys 1539
Db   91774 ACAGAAGAAACAGCCACCGAAATGTTGACGTCTGAGTTACAGACACAAGCTCACGTGATA 91833

Qy    1540 ArgLeuAspSerArgGluProAsnArgCysTrpThrLys---------------MetLys 1554
Db   91834 AAAGCTGACAGCCAGGACGCTGGACAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCC 91893

Qy    1555 AspAla-LysMetLysHisProValProGlnProArgGlu------------------- 1567
Db   91894 TCTGCACAGGATGAAACACCAATTACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTG 91953

Qy    1568 -AspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLeuProAr 1587
Db   91954 GGACAAGCACATTCTGATATTTCCAAAGACATGAGTGAAGCCTCAGAAAAGACCATGACT 92013

Qy    1587 gLeuGlnLeuLysAlaProVal 1594
Db   92014 GTTGAGGTAGAAGGTTCCACTG 92035
```

Search completed: December 13, 2002, 03:13:17
Job time : 9259 secs

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

```
Run on:         December 13, 2002, 00:16:41 ; Search time 619 Seconds
                                  (without alignments)
                                  5806.445 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters:      4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09902432/runat_12122002_141451_25790/app_query.fasta_1.1735
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09902432_@CGN_1_1_488_@runat_12122002_141451_25790 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      N_Geneseq_101002:*
                1:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
                2:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
                3:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
                4:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
                5:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
                6:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
                7:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
                8:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
                9:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
                10:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
                11:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
                12:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
                13:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
                14:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
                15:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
                16:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
                17:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
                18:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
                19:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
                20:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
                21:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
                22:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
                23:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
                24:  /SIDS2/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	8041	99.6	5200	18	AAV02302	Rat tumour suppres
	2	6001	74.3	5074	18	AAV02301	Rat tumour suppres
	3	4340.5	53.8	6614	24	ABN59861	Novel human coding
	4	4300.5	53.3	6608	24	ABN96942	Gene #3440 used to
	5	4300.5	53.3	6608	24	ABL61938	Colon adenocarcino
	6	4281	53.0	6605	19	AAV23545	Human gravin polyp
	7	4281	53.0	6605	21	AAA74903	Human gravin codin
	8	4226.5	52.4	6886	23	AAS85205	DNA encoding novel
	9	642	8.0	8307	23	ABL23161	Drosophila melanog
	10	639.5	7.9	11597	23	ABL23160	Drosophila melanog
	11	565.5	7.0	18506	23	ABL05592	Drosophila melanog
	12	557.5	6.9	20448	23	ABL19989	Drosophila melanog
c	13	537	6.7	24971	23	ABL19988	Drosophila melanog
	14	530.5	6.6	24789	23	ABL28640	Drosophila melanog
	15	492	6.1	7568	23	AAS80823	DNA encoding novel
	16	492	6.1	9220	23	ABV24414	Human prostate exp
	17	492	6.1	9220	23	ABV25265	Human prostate exp
	18	492	6.1	9416	24	ABK83800	Human cDNA differe
	19	490.5	6.1	10242	23	ABL20787	Drosophila melanog
	20	462	5.7	5361	18	AAT78868	P. falciparum live
	21	462	5.7	5529	24	ABK50883	DNA encoding Plasm
	22	462	5.7	6152	18	AAT78867	P. falciparum live
	23	454	5.6	6919	23	ABL05593	Drosophila melanog
	24	453.5	5.6	4407	23	ABL12349	Drosophila melanog
c	25	451	5.6	6561	23	ABL12348	Drosophila melanog
c	26	450.5	5.6	20978	23	ABL20786	Drosophila melanog
	27	448	5.5	9953	23	ABL21371	Drosophila melanog
	28	440.5	5.5	5439	23	ABL02411	Drosophila melanog
	29	432	5.4	9373	23	ABL07021	Drosophila melanog
	30	431.5	5.3	7997	22	ABA47417	Human breast cell
	31	431.5	5.3	7997	22	ABA65305	Human foetal liver
	32	431.5	5.3	7997	22	ABA32406	Probe #10872 for g
	33	431.5	5.3	7997	22	AAK13718	Human brain expres
	34	431.5	5.3	7997	22	AAK39460	Human bone marrow
	35	431.5	5.3	7997	22	AAI20270	Probe #10203 for g
	36	431.5	5.3	7997	22	AAI45472	Probe #14158 used
	37	431.5	5.3	7997	22	AAI05974	Probe #5965 used t
	38	431.5	5.3	7997	24	ABS13549	Human genome-deriv
	39	431.5	5.3	11680	21	AAC76700	Human ORFX ORF2255
	40	430	5.3	12505	23	ABL07020	Drosophila melanog
	41	429	5.3	8922	23	ABL04053	Drosophila melanog
c	42	429	5.3	15913	23	ABL04052	Drosophila melanog
	43	428.5	5.3	6744	13	AAQ29471	Extracellular fact
	44	427	5.3	15231	24	ABK63598	Rat sequence diffe
	45	422	5.2	10253	24	ABL60871	Human cytoskeleton

ALIGNMENTS

```
RESULT 1
AAV02302
ID   AAV02302 standard; cDNA; 5200 BP.
XX
AC   AAV02302;
XX
DT   20-JUL-1998  (first entry)
XX
DE   Rat tumour suppressor gene SSeCKS.
XX
KW   SSeCKS; tumour suppressor gene; rat; protein kinase C; mitosis;
KW   cancer; malignancy; cell proliferation; Alzheimer's disease;
KW   therapy; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..4791
```

```
FT                   /*tag=  a
XX
PN   WO9740059-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;   97WO-US06830.
XX
PR   18-JUN-1996;   96US-0665401.
PR   19-APR-1996;   96US-0635121.
XX
PA   (GELM/) GELMAN I.
PA   (JAKE/) JAKEN S.
XX
PI   Gelman I,  Jaken S;
XX
DR   WPI; 1997-535770/49.
DR   P-PSDB; AAW31347.
XX
PT   Tumour suppressor gene SSeCKS - used as a mitotic regulator, and
PT   inhibitor of malignant phenotype
XX
PS   Claim 1; Fig 11A-L; 162pp; English.
XX
CC   This rat cDNA sequence codes for the full-length form of the novel
CC   tumour suppressor protein SSeCKs (see AAW31347).  The SSeCKs gene
CC   product is a substrate of protein kinase C and acts as a negative
CC   regulator of mitosis and as an inhibitor of the transformed
CC   phenotype.  A partial SSeCKS clone (see AAV02301) was identified by
CC   searching for cDNAs whose abundance was low in NIH 3T3 cells and
CC   decreased following the expression of the activated oncogene v-src.
CC   A 5'RACE product was spliced to this truncated sequence to construct
CC   the full-length SSeCKS cDNA sequence.  This full-length SSeCKS
CC   nucleic acid sequence, as well as homologous and hybridising nucleic
CC   acids are claimed, as are vectors comprising such nucleic acids,
CC   encoded proteins, host cells and methods of inhibiting the
CC   expression of a transformed phenotype in a host cell by introducing
CC   the nucleic acid.  Introduction of a SSeCKS nucleic acid or gene
CC   product into a host cell inhibits mitosis of the host cell,
CC   allowing the treatment of diseases associated with disorders of
CC   proliferation and/or with the expression of a malignant phenotype.
CC   SSeCKS can also be used to treat or identify disorders of
CC   cytoskeletal structure and cellular architecture (such as
CC   Alzheimer's disease), and may be a marker for aberrancies in
CC   fertility and/or nervous system development.
XX
SQ   Sequence 5200 BP; 1612 A; 1201 C; 1563 G; 824 T; 0 other;

Alignment Scores:
Pred. No.:              0             Length:         5200
Score:                  8041.00       Matches:        1592
Percent Similarity:     99.75%        Conservative:   0
Best Local Similarity:  99.75%        Mismatches:     4
Query Match:            99.60%        Indels:         0
DB:                     18            Gaps:           0

US-09-902-432-4 (1-1596) x AAV02302 (1-5200)

Qy        1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGAGCGACACG 60

Qy       21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CCGAGCGAGCTGGTGCTCAGTGGCCATGGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 120

Qy       41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 180

Qy       61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 240

Qy       81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 300

Qy      101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 360

Qy      121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      361 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 420

Qy      141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      421 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 480

Qy      161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
            |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db      481 GTTGGTTTTATTTTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 540

Qy      181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      541 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 600

Qy      201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      601 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 660

Qy      221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      661 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 720

Qy      241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      721 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 780

Qy      261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      781 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 840

Qy      281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
            |||||||||   ||||||||||||||||||||||||||||||||||||||||||||||||
Db      841 ACATCTTCCTCCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 900

Qy      301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      901 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 960

Qy      321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      961 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1020

Qy      341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1021 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1080

Qy      361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1081 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1140

Qy      381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1141 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1200

Qy      401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1201 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1260

Qy      421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1261 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1320
```

```
Qy    441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1380

Qy    461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1440

Qy    481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1500

Qy    501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAG 1560

Qy    521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1620

Qy    541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1680

Qy    561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1740

Qy    581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1800

Qy    601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1860

Qy    621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1920

Qy    641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 1980

Qy    661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2040

Qy    681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2100

Qy    701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2160

Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
          ||||||||||||||||||||||||||||||||||||||||||||   |||||||||
Db   2161 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCTTTCCGAAGGG 2220

Qy    741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2280

Qy    761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2340

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2400

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2460

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2520

Qy    841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2580

Qy    861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2640

Qy    881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2700

Qy    901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2760

Qy    921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2761 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2820

Qy    941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2821 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2880

Qy    961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2881 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2940

Qy    981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2941 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3000

Qy   1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3001 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3060

Qy   1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3061 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3120

Qy   1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3121 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3180

Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3181 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3240

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3241 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3300

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3301 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3360

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3361 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3420

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3421 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3480

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3481 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3540

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3541 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3600

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3601 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3660

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3661 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3720

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3721 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3780

Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3781 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3840

Qy   1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3841 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3900

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3901 AAGGAGAAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3960

Qy   1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3961 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4020

Qy   1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4021 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4080

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4081 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4140

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4141 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4200

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
          |||||     ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4201 GTAGGTGACCACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4260

Qy   1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4261 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4320

Qy   1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4321 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4380

Qy   1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4381 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4440

Qy   1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4441 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4500

Qy   1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4501 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4560

Qy   1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4561 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4620

Qy   1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4621 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4680

Qy   1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4681 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4740

Qy   1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   4741 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4788
```

```
RESULT 2
AAV02301
ID   AAV02301 standard; cDNA; 5074 BP.
XX
AC   AAV02301;
XX
DT   20-JUL-1998  (first entry)
XX
DE   Rat tumour suppressor gene SSeCKS (truncated form).
XX
KW   SSeCKS; tumour suppressor gene; rat; protein kinase C; mitosis;
KW   cancer; malignancy; cell proliferation; Alzheimer's disease;
KW   therapy; ss.
XX
OS   Rattus sp.
XX
FH   Key             Location/Qualifiers
FT   CDS             176..4216
FT                   /*tag=  a
FT                   /transl_except= (pos:389..391, aa:Glu)
FT                   /transl_except= (pos:1661..1663, aa:Lys)
FT                   /transl_except= (pos:1668..1670, aa:Lys)
FT                   /transl_except= (pos:2444..2446, aa:Arg)
FT                   /transl_except= (pos:2528..2530, aa:Glu)
FT   polyA_signal    2528..2530
FT                   /*tag=  b
FT   polyA_signal    5056..5061
FT                   /*tag=  c
XX
PN   WO9740059-A1.
XX
PD   30-OCT-1997.
XX
PF   18-APR-1997;   97WO-US06830.
XX
PR   18-JUN-1996;   96US-0665401.
PR   19-APR-1996;   96US-0635121.
XX
PA   (GELM/) GELMAN I.
PA   (JAKE/) JAKEN S.
XX
PI   Gelman I,  Jaken S;
XX
DR   WPI; 1997-535770/49.
DR   P-PSDB; AAW31346.
XX
PT   Tumour suppressor gene SSeCKS - used as a mitotic regulator, and
PT   inhibitor of malignant phenotype
XX
PS   Example 6; Fig 3A-H; 162pp; English.
XX
CC   This rat cDNA sequence codes for an active truncated form of the
CC   novel tumour suppressor protein SSeCKs (see AAW31346).  The SSeCKs
CC   gene product is a substrate of protein kinase C and acts as a
CC   negative regulator of mitosis and as an inhibitor of the transformed
CC   phenotype.  The SSeCKS clone was identified by searching for cDNAs
CC   whose abundance was low in NIH 3T3 cells and decreased following the
CC   expression of the activated oncogene v-src.  A 5'RACE product was
CC   spliced to the truncated sequence to construct a full-length
CC   SSeCKS cDNA (see AAV02302).  This full-length SSeCKS nucleic acid
```

```
CC   sequence, as well as homologous and hybridising nucleic acids, are
CC   claimed, as are vectors comprising such nucleic acids, host cells
CC   and methods of inhibiting the expression of a transformed phenotype
CC   in a host cell by introducing such nucleic acids.  Introduction of
CC   a SSeCKS nucleic acid or gene product into a host cell inhibits
CC   mitosis of the host cell, allowing the treatment of diseases
CC   associated with disorders of proliferation and/or with the
CC   expression of a malignant phenotype.  SSeCKS can also be used to
CC   treat or identify disorders of cytoskeletal structure and cellular
CC   architecture (such as Alzheimer's disease), and may be a marker
CC   for aberrancies in fertility and/or nervous system development.
XX
SQ   Sequence 5074 BP; 1501 A; 1151 C; 1441 G; 981 T; 0 other;

Alignment Scores:
Pred. No.:              3.15e-300     Length:         5074
Score:                  6001.00       Matches:        1231
Percent Similarity:     97.33%        Conservative:   8
Best Local Similarity:  96.70%        Mismatches:     27
Query Match:            74.33%        Indels:         14
DB:                     18            Gaps:           4

US-09-902-432-4 (1-1596) x AAV02301 (1-5074)

Qy     329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 61

Qy     349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 121

Qy     369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 181

Qy     389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 241

Qy     409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 301

Qy     429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 361

Qy     449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 421

Qy     469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 481

Qy     489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 541

Qy     509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 601

Qy     529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 661

Qy     549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 721

Qy     569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588
           |||||||||||||||||  ||||||||||||||||||||||||||||||||   ||||||
Db     722 GGGCCGCTGGAAGCACCCAG-GATGGGGAAGCTGAGGAAGGAACTACTTCGT--GGAGAG 778

Qy     589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608
           |||||||||:::  ||||||||||||||||||||||||||||||||||||||||||||||
Db     779 AAGAAGAGGAAGG--ATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 836

Qy     609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628
           |||  |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     837 CGGTC-CGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 895

Qy     629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     896 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 955

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 1015

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATATAAGA 1075

Qy     689 GlyGlyProArgThrLeuGly-GlyAspSerHisArgAlaGluGluAlaSerLysAspLy 708
              |||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db    1076 ---GGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAA 1132

Qy     708 sGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySe 728
           ||||||  ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1133 AGAAGCCGA-ACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAG 1191

Qy     728 rSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSe 748
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1192 TTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTC 1251

Qy     748 rPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla-- 767
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1252 ATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA-AG 1310

Qy     768 ----GluAspSerSerVal---GluGlnLeuSerThrGluIleGluProSerArgGluGl 785
               :::|||||||||       :::||||||||||||||||||||||    ||||||||
Db    1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy     785 uSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLy 805
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy     805 s-GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProA 825
           | |||||||||||||||||||||||||||||||||||||||||||||||||:::|||:
Db    1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy     825 snValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluA 845
           ::|||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db    1491 ATGTCCCAGCAGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy     845 laGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865
           ||||||||||| |||||||||          |||  |||   ||||||||||||||||
Db    1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy     865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy     885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy     905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluG 925
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy     925 luThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValT 945
           ||||||||||||||||||||||||||||||||||:::|||||||||||||||||||||||
```

```
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTCAAGATGCCCATGACGACATGGTCA 1847

Qy    945 hrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuA 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy    965 rgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaV 985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy    985 alSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValG 1005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   1005 luSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaV 1025
          ||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   1025 alAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrG 1045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   1045 lySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerG 1065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   1065 luLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuA 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   1085 laGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrA 1105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   1105 laAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnA 1125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   1125 laValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProL 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   1145 euAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspS 1165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   1165 erLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaG 1185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   1185 lnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluG 1205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   1205 luProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProV 1225
          ||||||||||||||||||||||||||||||||||||||     |||||||||||||||||
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   1225 alLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysV 1245
          ||       |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   1245 alLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspV 1265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   1265 alThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluV 1285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   1285 alGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgG 1305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   1305 luThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisG 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   1325 luGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysA 1345
          ||        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   1345 laLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleG 1365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   1365 luValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysV 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   1385 alIleGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisL 1405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   1405 euLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluA 1425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   1425 spThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProA 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   1445 laProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluA 1465
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   1465 rgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrA 1485
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   1485 laIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnL 1505
          |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   1505 ysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaP 1525
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   1525 roGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgG 1545
          |||||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   1545 luProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnP 1565
          ||||||||||||||||||||||||||||||    ||||||||||||||||||||||||||
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   1565 roArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysL 1585
          ||||||||||||||||||||||||||||||||||||||||||          ||||||||
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGC--TCAGCTCGGAAATGCT 3765

Qy   1585 euProArgLeuGlnLeuLysAlaProValSerLys 1596
          ||||||||||||||||||||||||||||||||||||
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 3800
```

```
RESULT 3
ABN59861
ID   ABN59861 standard; cDNA; 6614 BP.
XX
AC   ABN59861;
XX
DT   28-JUN-2002  (first entry)
XX
DE   Novel human coding sequence SEQ ID NO: 272.
XX
KW   Human; antianaemic; vulnerary; antiinflammatory; immunomodulator;
KW   antiinfertility; cerebroprotective; cytostatic; rheumatic; gene therapy;
```

```
KW   neuroprotective; antiparkinsonian; protein therapy; EST;
KW   expressed sequence tag; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200222660-A2.
XX
PD   21-MAR-2002.
XX
PF   10-SEP-2001; 2001WO-US26015.
XX
PR   11-SEP-2000; 2000US-0659671.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Zhou P,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Wehrman T,  Drmanac RT;
XX
DR   WPI; 2002-292408/33.
DR   P-PSDB; ABB97448.
XX
PT   An isolated polynucleotide for treating diseases associated with its
PT   encoded polypeptide such as cancer and multiple sclerosis -
XX
PS   Claim 1; SEQ ID NO 272; 509pp; English.
XX
CC   The present invention provides the protein and coding sequences of 444
CC   novel human proteins. These were isolated from expressed sequences tags
CC   (ESTs). They can be used to stimulate cell growth, to regulate
CC   haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth
CC   e.g. in burn treatment, to regulate the immune system e.g. to treat
CC   multiple sclerosis, to regulate activin or inhibin e.g. to treat
CC   infertility, to regulate haemostasis or thrombolysis e.g. to treat
CC   stroke and cancer, to screen for drugs, to treat inflammatory conditions
CC   e.g. rheumatoid arthritis, and to treat nervous system disorders e.g.
CC   Parkinson's disease. The present sequence is a coding sequence of the
CC   invention.
XX
SQ   Sequence 6614 BP; 2094 A; 1420 C; 1827 G; 1273 T; 0 other;

Alignment Scores:
Pred. No.:              1.48e-214      Length:         6614
Score:                  4340.50        Matches:        970
Percent Similarity:     69.13%         Conservative:   190
Best Local Similarity:  57.81%         Mismatches:     408
Query Match:            53.77%         Indels:         110
DB:                     24             Gaps:           33

US-09-902-432-4 (1-1596) x ABN59861 (1-6614)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln------------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACAGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGACACCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 671

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    672 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 731

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    732 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 791

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    792 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 842

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    843 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 896

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    897 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 956

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    957 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1013

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1014 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1073

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1074 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1133

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1134 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1193

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1194 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1253

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1254 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1313

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1314 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1373

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1374 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1433

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1434 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1484

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1485 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1529

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1530 TTGGTTGAAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1589
```

```
Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1590 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1649

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1650 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1709

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1710 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1769

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1770 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1826

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1827 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1886

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1887 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1946

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1947 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2006

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2007 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2066

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2067 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2126

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2127 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2186

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2187 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2246

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2247 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2306

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2307 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2366

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2367 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGCAGGGAAGTTCCTCCCCGGAG 2426

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2427 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2486

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2487 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2546

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2547 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2606

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2607 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGAAACAAGAACAA 2666

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2667 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2726

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2727 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2786

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2787 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2846

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2847 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2906

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2907 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2966

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2967 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3026

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3027 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3086

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3087 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3146

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3147 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3206

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3207 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3266

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3267 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3326

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3327 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3386

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3387 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3428

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3429 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3488

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3489 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3548

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3549 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3608

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3609 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3668

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3669 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3728

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
```

```
Db    3729 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3788

Qy    1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db    3789 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3848

Qy    1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db    3849 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3905

Qy    1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db    3906 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3965

Qy    1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db    3966 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4019

Qy    1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db    4020 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4079

Qy    1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db    4080 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4139

Qy    1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db    4140 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4199

Qy    1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db    4200 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4259

Qy    1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db    4260 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4319

Qy    1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db    4320 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4379

Qy    1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db    4380 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4439

Qy    1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db    4440 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4499

Qy    1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db    4500 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4559

Qy    1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db    4560 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4619

Qy    1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db    4620 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4679

Qy    1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db    4680 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4739

Qy    1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db    4740 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGGATTTA 4793

Qy    1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
Db    4794 GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAAAACATC 4853

Qy    1511 IleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThrHisAla 1529
Db    4854 ATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATGTTGACG 4913

Qy    1530 TyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCys 1549
Db    4914 TCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGACAGGAA 4973

Qy    1550 TrpThrLys--------------MetLysAspAla-LysMetLysHisProValProGl 1564
Db    4974 ACGGAGAAAGAAGGAGAGGACCCTCAGGCCTCTGCACAGGATGAAACACCAATTACTTCA 5033

Qy    1564 nProArgGlu--------------------AspLeuGlnValLeuThrValLeuGluAl 1577
Db    5034 GCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCCAAAGAC 5093

Qy    1577 aTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db    5094 ATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5145
```

```
RESULT 4
ABN96942
ID   ABN96942 standard; DNA; 6608 BP.
XX
AC   ABN96942;
XX
DT   13-AUG-2002  (first entry)
XX
DE   Gene #3440 used to diagnose liver cancer.
XX
KW   Gene; liver cancer; ds; hepatocellular carcinoma; hepatotropic;
KW   metastatic liver tumour; cytostatic; expression profile; disease state;
KW   disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS   Homo sapiens.
XX
PN   WO200229103-A2.
XX
PD   11-APR-2002.
XX
PF   02-OCT-2001; 2001WO-US30589.
XX
PR   02-OCT-2000; 2000US-237054P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Horne D,  Alvares C,  Peres-Da-Silva S,  Vockley JG;
XX
DR   WPI; 2002-426119/45.
XX
PT   Diagnosing and detecting the progression of liver cancer,
PT   hepatocellular carcinoma or metastatic liver tumor in a patient,
PT   involves detecting the level of expression of two or more genes in a
PT   liver tissue sample  -
XX
PS   Claim 1; SEQ ID NO 3440; 298pp; English.
XX
CC   The invention relates to a novel method for diagnosing and detecting the
CC   progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC   tumour in a patient, and differentiating metastatic liver cancer from
CC   hepatocellular carcinoma in a patient, involving detecting the level of
CC   expression of two or more genes represented in ABN93503-ABN97455 in a
CC   tissue sample. The method of the invention has hepatotropic, and
CC   cytostatic activity. The method is useful for diagnosing and detecting
CC   the progression of liver cancer, hepatocellular carcinoma and metastatic
CC   liver carcinoma in a patient. The method is useful for identifying
CC   expression profiles which serve as useful diagnostic markers as well as
CC   markers that can be used to monitor disease states, disease progression,
CC   drug toxicity, drug efficacy and drug metabolism.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
```

```
XX
SQ   Sequence 6608 BP; 2091 A; 1418 C; 1827 G; 1272 T; 0 other;

Alignment Scores:
Pred. No.:              1.7e-212     Length:         6608
Score:                  4300.50      Matches:        967
Percent Similarity:     68.81%       Conservative:   189
Best Local Similarity:  57.56%       Mismatches:     408
Query Match:            53.27%       Indels:         116
DB:                     24           Gaps:           36

US-09-902-432-4 (1-1596) x ABN96942 (1-6608)

Qy        1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
            ||||||||||||||||||||||||||||||||||||||||||   |||  |||||
Db      192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy       20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
            ||||||:::|||          |||||  ||||||:::|||||||||:::
Db      252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy       39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
                     |||||||||  |||         ||||||||||||||||  ||||||||||||
Db      312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy       57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
            ||||||||::::::|||||||||||||||||     ::::::   :::||||||    :::
Db      372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy       77 GlnGluGlyGln------------------------------------------GluGlu 82
                ||||||                                          ||||||
Db      432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy       83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
            ||||||:::      :::|||||||||||||:::||||||||||||   |||:::|||
Db      492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy      103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
              :::|||||||||       |||     |||   |||||||||   ||||||||||||
Db      552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy      123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
            |||            ||||||||||||||:::|||||:::|||:::||||||:::
Db      612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy      143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
            ||||||    ||||||||||||||||||:::||||||||||||||||||||||||||||||
Db      669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy      163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
            |||||||||||||||||||||||||||      ||||||   ||||||||||||||||||||
Db      729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy      183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
            ||||||||||||||||||||||||||||                 |||||||||||||:::
Db      789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy      203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
            ||||||:::                |||||:::||||||||||||||||||   ||||||||
Db      840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy      223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
            ||||||||    |||    |||||||||:::||||||||        ||||||
Db      894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy      243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
            |||||||||    ||||||          ||||||   |||:::|||||||||||||||||||
Db      954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy      263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
            ||||||||:::||||||   ||||||||:::||||||||||    ||||||||||    |||
Db     1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy      283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
            :::||||||||||||||||||    ||||||||||||||||||||||||||||||||:::
Db     1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy      303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
            |||   |||||||||::::::|||      :::|||||||:::|||||||||||||    ||||||
Db     1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy      323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
            ||||||    ||||||      |||    |||    |||||||||:::
Db     1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy      338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
                        ||||||||||||||            :::   ||||||||||:::|||
Db     1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy      354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
            ||||||||||||||||||   |||:::||||||          :::        |||||||||
Db     1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy      374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
            |||   |||||||||||||||||||||||||||||||||||:::|||   ||||||   |||
Db     1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy      393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
            ||||||||||||||||||||||||||||||||||:::         |||||||||
Db     1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy      413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
                           ||||||   |||          |||:::|||    |||:::
Db     1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy      433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
            |||                   ||||||||||||||||||||:::||||||:::|||
Db     1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy      453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
            :::|||    |||||||||||||    |||    ||||||         ||||||||||:::|||
Db     1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy      473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
            |||   |||    |||         |||||||:::||||||||||||||||||||||||||||
Db     1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy      493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
            ||||||:::||||||||||||||||||||||||||||||:::|||:::||||||||
Db     1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy      513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
            ||||||||||||||||||||||||||||||      |||||||||||||||    |||
Db     1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy      533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
            |||:::      :::      :::||||||:::|||      :::|||||||||||||||||
Db     1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy      553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
            ||||||||||||||||||||||||||||||||||||||||||||||||||||    |||||
Db     1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy      573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
                     ||||||||||||||||||||||||   ||||||||||||||||||||||||||
Db     1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy      593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
            |||:::||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy      613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
            ||||||||||||||||||||:::|||||||:::||||||||||||||||||||||||||||
Db     2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy      633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
```

```
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3845

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3846 AGGCCTCCAGCA---CCTTCCAGTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3902

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3903 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3962

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3963 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4016

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4017 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4076

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4077 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4136

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4137 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4196

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4197 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4256

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4257 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4316
```

```
Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
          |||:::::::::|||       :::       |||   :::   ||||||   ||||||
Db   4317 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4376

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
          |||            |||:::::|||   |||   :::::||||||||||||   :::
Db   4377 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4436

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
             :::   |||   ||||||   |||   |||:::   :::      ||||||
Db   4437 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4496

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
          |||   ||||||||||||:::   |||      ||||||||||||:::      ||||||
Db   4497 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4556

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
          ||||||                :::                  ||||||   ||||||   |||
Db   4557 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4616

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
          |||:::   ||||||:::|||   |||:::||||||:::            :::   |||
Db   4617 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4676

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluAsp 1470
                   ||||||:::||||||   :::   |||   :::   :::|||:::|||   |||
Db   4677 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4736

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
          :::      |||         :::         |||   |||   |||||||||   :::
Db   4737 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4787

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
                ||||||||         |||||||||||||||:::||||||:::|||:::|||
Db   4788 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4841

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
          |||:::||||||||||||||||||||||||   ||||||   |||||||||   |||
Db   4842 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4901

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
                   :::   |||||||||            :::   ||||||::::::
Db   4902 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4961

Qy   1548 ArgCysTrpThrLys--------------MetLysAspAla-LysMetLysHisProVa 1562
          :::         |||                   |||   :::|||||||||   ::
Db   4962 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5021

Qy   1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
          :   ||||||:::                   |||      :::|||
Db   5022 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5081

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
             :::         :::|||:::|||            |||:::   |||   |||:::
Db   5082 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5139
```

RESULT 5
ABL61938
ID   ABL61938 standard; DNA; 6608 BP.
XX
AC   ABL61938;
XX
DT   15-MAY-2002 (first entry)
XX
DE   Colon adenocarcinoma related gene sequence SEQ ID NO:275.
XX
KW   Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
KW   stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
KW   cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;
KW   gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO200194629-A2.
XX
PD   13-DEC-2001.
XX
PF   30-MAY-2001; 2001WO-US10838.
XX
PR   05-JUN-2000; 2000US-209473P.
PR   05-JUN-2000; 2000US-209531P.
PR   18-SEP-2000; 2000US-233133P.
PR   18-SEP-2000; 2000US-233617P.
PR   20-SEP-2000; 2000US-234009P.
PR   20-SEP-2000; 2000US-234034P.
PR   20-SEP-2000; 2000US-234052P.
PR   22-SEP-2000; 2000US-234509P.
PR   22-SEP-2000; 2000US-234567P.
PR   25-SEP-2000; 2000US-234923P.
PR   25-SEP-2000; 2000US-234924P.
PR   25-SEP-2000; 2000US-235077P.
PR   25-SEP-2000; 2000US-235082P.
PR   25-SEP-2000; 2000US-235134P.
PR   25-SEP-2000; 2000US-235280P.
PR   26-SEP-2000; 2000US-235637P.
PR   26-SEP-2000; 2000US-235638P.
PR   27-SEP-2000; 2000US-235711P.
PR   27-SEP-2000; 2000US-235720P.
PR   27-SEP-2000; 2000US-235840P.
PR   27-SEP-2000; 2000US-235863P.
PR   28-SEP-2000; 2000US-236028P.
PR   28-SEP-2000; 2000US-236032P.
PR   28-SEP-2000; 2000US-236033P.
PR   28-SEP-2000; 2000US-236034P.
PR   28-SEP-2000; 2000US-236109P.
PR   28-SEP-2000; 2000US-236111P.
PR   29-SEP-2000; 2000US-236842P.
PR   29-SEP-2000; 2000US-236891P.
PR   02-OCT-2000; 2000US-237172P.
PR   02-OCT-2000; 2000US-237173P.
PR   02-OCT-2000; 2000US-237278P.
PR   02-OCT-2000; 2000US-237294P.
PR   02-OCT-2000; 2000US-237295P.
PR   02-OCT-2000; 2000US-237316P.
PR   03-OCT-2000; 2000US-237425P.
PR   03-OCT-2000; 2000US-237598P.
PR   03-OCT-2000; 2000US-237604P.
PR   03-OCT-2000; 2000US-237606P.
PR   03-OCT-2000; 2000US-237608P.
PR   01-NOV-2000; 2000US-244867P.
PR   01-NOV-2000; 2000US-245084P.
XX
PA   (AVAL-) AVALON PHARM.
XX
PI   Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;
PI   Soppet DR, Weaver Z;
XX
DR   WPI; 2002-188264/24.
XX
PT   Screening for anti-neoplastic agent involves exposing cells to a
PT   chemical agent to be tested for anti-neoplastic activity, and
PT   determining a change in expression of a gene of a signature gene set -
XX
PS   Claim 1; SEQ ID 275; 44pp; English.
XX
CC   The present invention describes a method (M1) for screening for an
CC   anti-neoplastic agent. The method involves exposing cells to a chemical
CC   agent to be tested for anti-neoplastic activity, determining a change in
CC   expression of at least one gene (I) of a signature gene set, where (I)
CC   comprises a sequence (S) selected from 8447 sequences (given in ABL61664
CC   to ABL70110), or is at least 95% identical to (S), where a change in
CC   expression is indicative of anti-neoplastic activity. (I) has cytostatic
CC   activity and can be used in gene therapy. M1 can be used for screening
CC   an anti-neoplastic agent, and can be used for producing a product which
CC   is the data collected with respect to the anti-neoplastic agent as a

```
CC   result of M1, and the data is sufficient to convey the chemical
CC   structure and/or properties of the agent. M1 can be used in the
CC   treatment of cancer such as colon, breast, stomach, lung, thyroid,
CC   oesophageal, ovarian, kidney, prostate or pancreatic cancer,
CC   adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,
CC   infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine
CC   carcinoma, papillary carcinoma and Wilm's tumour.
XX
SQ   Sequence 6608 BP; 2091 A; 1418 C; 1827 G; 1272 T; 0 other;

Alignment Scores:
Pred. No.:              1.7e-212      Length:         6608
Score:                  4300.50       Matches:        967
Percent Similarity:     68.81%        Conservative:   189
Best Local Similarity:  57.56%        Mismatches:     408
Query Match:            53.27%        Indels:         116
DB:                     24            Gaps:           36

US-09-902-432-4 (1-1596) x ABL61938 (1-6608)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln------------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
```

```
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy    613 ProSerGluSerAspLysGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA-----------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3845

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3846 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3902

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3903 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3962

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3963 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4016

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4017 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4076

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4077 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4136

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4137 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4196
```

```
Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4197 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4256

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4257 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4316

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4317 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4376

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4377 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4436

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4437 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4496

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4497 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4556

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4557 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4616

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4617 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4676

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4677 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4736

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4737 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4787

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
Db   4788 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4841

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
Db   4842 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4901

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
Db   4902 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4961

Qy   1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
Db   4962 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5021

Qy   1562 lProGlnProArgGlu---------------------AspLeuGlnValLeuThrValLe 1575
Db   5022 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5081

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   5082 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5139
```

```
RESULT 6
AAV23545
ID   AAV23545 standard; cDNA; 6605 BP.
XX
AC   AAV23545;
XX
DT   13-JUL-1998  (first entry)
XX
DE   Human gravin polypeptide coding sequence.
XX
KW   Gravin; kinase anchoring protein; type II regulatory subunit; PKA; PKC;
KW   cAMP-dependent protein kinase; protein kinase C; autoimmune disease;
KW   Myasthenia gravis; nicotinic acetylcholine receptor; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             192..5534
FT                   /*tag=  a
XX
PN   US5741890-A.
XX
PD   21-APR-1998.
XX
PF   19-DEC-1996;   96US-0769309.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Nauert JB,  Scott JD;
XX
DR   WPI; 1998-260552/23.
DR   P-PSDB; AAW53863.
XX
PT   New polypeptide fragments of protein kinase binding protein gravin -
PT   are useful for the study of modulation of action between gravin and
PT   protein kinase(s)
XX
PS   Example 1; Column 19-32; 32pp; English.
XX
CC   This sequence encodes the human gravin polypeptide, which represents
CC   a polypeptide of the invention. The polypeptides are fragments capable of
CC   binding to type II regulatory subunit of cAMP-dependent protein kinase
CC   (PKA). Gravin is a kinase anchoring protein that binds to type II
CC   regulatory subunits of PKA and protein kinase C (PKC). Gravin is also an
CC   antigen of the autoimmune disease Myasthenia gravis (MG), where a patient
CC   develops antibodies against their own nicotinic acetylcholine receptors.
CC   The polypeptides are useful for providing analogues of gravin in the
CC   study of the modulation (e.g. blocking, inhibiting and stimulating) of
CC   interactions between gravin and kinase. The peptides are involved in the
CC   modulation of gravin-kinase interactions.
XX
SQ   Sequence 6605 BP; 2091 A; 1416 C; 1826 G; 1272 T; 0 other;

Alignment Scores:
Pred. No.:              1.72e-211      Length:         6605
Score:                  4281.00        Matches:        965
Percent Similarity:     68.69%         Conservative:   189
Best Local Similarity:  57.44%         Mismatches:     409
Query Match:            53.03%         Indels:         117
DB:                     19             Gaps:           37

US-09-902-432-4 (1-1596) x AAV23545 (1-6605)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln-----------------------------------------GluGlu 82
```

```
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491
Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551
Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611
Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668
Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728
Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788
Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839
Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893
Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953
Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010
Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070
Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130
Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190
Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250
Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310
Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370
Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430
Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481
Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432

Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526
Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586
Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646
Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706
Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766
Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823
Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883
Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943
Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003
Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063
Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123
Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183
Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243
Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303
Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363
Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGAAGTTCCTCCCCGGAG 2423
Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483
Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543
Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603
```

```
Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTGCATCTGGTA---CCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3842

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3843 AGGCCTCCAGCA---CCTTCCAGTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3899

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3900 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3959

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3960 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4013

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4014 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4073

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4074 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4133

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4134 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4193

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4194 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4253

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4254 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4313

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4314 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4373

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4374 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4433

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4434 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4493

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4494 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4553

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4554 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4613

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4614 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4673

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4674 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4733

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
```

```
            :::       |||       :::          |||   |||   ||||||||   :::
Db    4734 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4784

Qy    1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
                 ||||||||    ||||||||||||||:::||||||:::|||:::|||
Db    4785 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4838

Qy    1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
           |||:::|||||||||||||||||||||||   ||||||   |||||||||   |||
Db    4839 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4898

Qy    1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
                 :::    ||||||||              :::    ||||||::::::
Db    4899 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4958

Qy    1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
           :::          |||                       ||| :::|||||||||    ::
Db    4959 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5018

Qy    1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
            :    ||||||:::                      |||        :::|||
Db    5019 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5078

Qy    1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
             :::             :::|||:::|||          |||:::   |||   |||:::
Db    5079 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5136
```

RESULT 7
AAA74903

```
ID   AAA74903 standard; cDNA; 6605 BP.
XX
AC   AAA74903;
XX
DT   26-JAN-2001  (first entry)
XX
DE   Human gravin coding sequence.
XX
KW   Human; gravin; PKA RII binding site; myasthenia gravis;
KW   kinase anchoring protein; cAMP dependent protein kinase; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             192..5534
FT                   /*tag=  a
FT                   /product= "gravin"
XX
PN   US6090929-A.
XX
PD   18-JUL-2000.
XX
PF   19-DEC-1997;   97US-0994570.
XX
PR   19-DEC-1996;   96US-0769309.
XX
PA   (UYOR-) UNIV OREGON HEALTH SCI.
XX
PI   Klauck TM,  Scott JD,  Nauert JB;
XX
DR   WPI; 2000-523763/47.
DR   P-PSDB; AAB15380.
XX
PT   Novel polynucleotides useful for detecting gravin in patients suffering
PT   from Myasthenia gravis encodes cAMP-dependent protein kinase-binding
PT   polypeptide and protein kinase C-binding polypeptide of gravin  -
XX
PS   Example 1; Column 23-36; 34pp; English.
XX
CC   The present sequence is the coding sequence for human gravin. Gravin is
CC   an A-kinase anchoring protein (AKAP) which is involved in the
CC   localisation of cAMP dependent protein kinase A (PKA) via interactions
CC   between the RII binding region and the PKA regulatory subunit RII. Gravin
CC   is also an antigen found in myasthenia gravis sufferers, and it is
CC   thought that antibodies to it may be useful in modulating the binding of
CC   PKA, and thus aid in the treatment of the disease. This sequence was
CC   isolated by first screening a human umbilical vein endothelial cell cDNA
CC   library with serum from a myasthenia gravis patient, and then searching a
CC   human heart cDNA library for sequences resembling the isolated sequence.
CC   This was done because the first sequence obtained was shown to be shorter
CC   than the full length coding sequence.
XX
SQ   Sequence 6605 BP; 2091 A; 1416 C; 1826 G; 1272 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              1.72e-211      Length:        6605
Score:                  4281.00        Matches:       965
Percent Similarity:     68.69%         Conservative:  189
Best Local Similarity:  57.44%         Mismatches:    409
Query Match:            53.03%         Indels:        117
DB:                     21             Gaps:          37
```

US-09-902-432-4 (1-1596) x AAA74903 (1-6605)

```
Qy       1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
           ||||||||||||||||||||||||||||||||||||||||||   |||   ||||||
Db     192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy      20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
           ||||||:::|||         ||||||   ||||||:::||||||||||||:::
Db     252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy      39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
                 |||||||||   |||      ||||||||||||||||||   ||||||||||||
Db     312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy      57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
           |||||||||::::::||||||||||||||||||      ::::::   :::||||||   :::
Db     372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy      77 GlnGluGlyGln------------------------------------------GluGlu 82
                 ||||||                                         ||||||
Db     432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy      83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
           ||||||:::         :::||||||||||||:::||||||||||||   |||:::|||
Db     492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy     103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
              :::|||||||||         |||         |||   |||||||||   ||||||||||||
Db     552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy     123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
           |||            ||||||||||||||:::||||||:::|||:::||||||:::
Db     612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy     143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
           ||||||   ||||||||||||||||||:::||||||||||||||||||||||||||||||
Db     669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy     163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
           ||||||||||||||||||||||||||||      ||||||   ||||||||||||||||||
Db     729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy     183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
           ||||||||||||||||||||||||||||         ||||||||||||||||||:::
Db     789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy     203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
           ||||||:::            ||||||:::||||||||||||||||||   |||||||||
Db     840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy     223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
           |||||||||   |||   |||||||||:::|||||||||         ||||||
Db     894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy     243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
```

```
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143
```

```
Qy     965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
              :::||||||  ||||||||||  ||||||||||||:::||||||||||
Db    3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy     985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy    1005 GluSerGlyValLeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
           |||  ||||||  |||  ||||||:::||||||:::||||||   :::||||||||
Db    3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy    1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr----------------- 1038
           ||||||:::|||||||||||||||||||||:::|||   |||
Db    3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy    1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
              |||   ||||||||                   :::   |||    ||||||
Db    3384 CTTCAGCCTGTGCAGAGAGCA-----------------GAGGCAGAAAGACCAGAAGAG 3425

Qy    1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
             :::|||               :::||||   ||||||:::  |||     |||
Db    3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy    1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
           ||||||      |||        ||||||          ||||||   |||||||||||
Db    3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy    1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
              |||      |||:::||||||       ::::::          |||   ||||||
Db    3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy    1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
           |||                  :::|||   |||::::::   ||||||||||||:::
Db    3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy    1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
           |||:::|||   ||||||    |||||||||||||:::|||||||||||||:::||||||
Db    3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy    1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
              |||            ||||||:::||||||    ::::::      :::
Db    3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy    1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
                       |||          :::   ||||||   ||||||   ||||||||||||
Db    3786 GTGCATCTGGTA---CCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3842

Qy    1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
              |||         |||::::::      ||||||||        |||:::
Db    3843 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3899

Qy    1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
                |||  ||||||   |||   :::|||::::::            |||   :::|||
Db    3900 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3959

Qy    1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
           :::|||||||||      |||     |||   |||        ||||||   |||
Db    3960 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4013

Qy    1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
                          |||   |||    :::    :::|||          :::
Db    4014 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4073

Qy    1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
           :::|||     :::|||            |||            |||   |||::::::
Db    4074 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4133

Qy    1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
           ::::::     |||:::||||||                |||          |||   |||
Db    4134 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4193

Qy    1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
           |||        ||||||:::||||||   |||   :::|||      |||:::||||||
Db    4194 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4253

Qy    1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
                 |||   ||||||||||         |||          :::     :::
Db    4254 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4313

Qy    1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
           |||:::::::::|||         :::         |||   :::   ||||||   ||||||
Db    4314 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4373

Qy    1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
           |||               |||::::::|||   |||   ::::::||||||||||||   :::
Db    4374 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4433

Qy    1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
              :::   |||   ||||||   |||   |||:::   :::      ||||||
Db    4434 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4493

Qy    1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
           |||   ||||||||||||:::   |||      ||||||||||||:::      ||||||
Db    4494 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4553

Qy    1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
           ||||||         :::            ||||||   ||||||   ||||||   |||
Db    4554 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4613

Qy    1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
           |||:::   ||||||:::|||   |||:::||||||:::            :::   |||
Db    4614 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4673

Qy    1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
                 ||||||:::||||||   :::   |||   :::   :::|||:::|||   |||
Db    4674 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4733

Qy    1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
           :::      |||      :::         |||   |||   |||||||||   :::
Db    4734 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4784

Qy    1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
           ||||||||||||         ||||||||||||:::||||||:::|||:::|||
Db    4785 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4838

Qy    1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
           |||:::||||||||||||||||||||||||   ||||||   |||||||||   |||
Db    4839 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4898

Qy    1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
                 :::   |||||||||         :::   ||||||::::::
Db    4899 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4958

Qy    1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
           :::         |||                    |||  :::||||||||||   ::
Db    4959 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5018

Qy    1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
            :   ||||||:::                       |||   :::|||
Db    5019 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5078

Qy    1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
            :::      :::|||:::|||               |||:::   |||   |||:::
Db    5079 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5136

RESULT 8
AAS85205
ID   AAS85205 standard; cDNA; 6886 BP.
XX
AC   AAS85205;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #21009.
```

```
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG21018.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity -
XX
PS   Claim 1; SEQ ID No 21009; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 6886 BP; 2142 A; 1483 C; 1908 G; 1352 T; 1 other;

Alignment Scores:
Pred. No.:              1.16e-208     Length:         6886
Score:                  4226.50       Matches:        960
Percent Similarity:     68.48%        Conservative:   187
Best Local Similarity:  57.31%        Mismatches:     407
Query Match:            52.35%        Indels:         121
DB:                     23            Gaps:           36

US-09-902-432-4 (1-1596) x AAS85205 (1-6886)

Qy        1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
              ||||||||||||||||||||||||||||||||||||||||||      |||   ||||||
Db      192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy       20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
              ||||||:::|||           ||||||    ||||||:::|||||||||:::
Db      252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy       39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
                    |||||||||   |||      |||||||||||||||||     ||||||||||||

Db      312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy       57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
              ||||||||::::::||||||||||||||||||      ::::::    :::||||||    :::
Db      372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy       77 GlnGluGlyGln-----------------------------------------GluGlu 82
                     ||||||                                          ||||||
Db      432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy       83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
              ||||||:::            |||            :::||||||||||||     ::::::|||
Db      492 GAAGTCATTGTCCGGCAGGTTGATCACATAGACTCTGAAGATGTGAGCGACAGAGACTCC 551

Qy      103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
                     ||||||          |||        |||    ||||||||     ||||||||||
Db      552 GATACAGAGATGACTACTAAGTCAGCGGTAGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy      123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
              ||||||    :::||||||||||||||||||:::||||||:::|||:::||||||:::
Db      612 GAGACACCCGATATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 671

Qy      143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
              ||||||    ||||||||||||||||||:::||||||||||||||||||||||||||||||
Db      672 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 731

Qy      163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
              ||||||||||||||||||||||||||||||    ||||||   ||||||||||||||||||||
Db      732 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 791

Qy      183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
              ||||||||||||||||||||||||||||            ||||||||||||||||||:::
Db      792 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 842

Qy      203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
              ||||||:::                ||||||:::||||||||||||||||||    ||||||||
Db      843 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 896

Qy      223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro-LeuGl 242
              |||||||||    |||    |||||||||:::|||||||||         ||||||      ||
Db      897 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 956

Qy      242 nAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGl 262
              |||||||||    ||||||      ||||||      |||:::||||||||||||||||||||
Db      957 AGCGGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGA 1013

Qy      262 uLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSe 282
              |||||||||:::||||||    |||||||||:::|||||||||     |||||||||    ||
Db     1014 AAAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATC 1073

Qy      282 rSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLy 302
              |:::||||||||||||||||||    ||||||||||||||||||||||||||||||||::
Db     1074 AACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAG 1133

Qy      302 sLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLy 322
              :|||    ||||||||||::::::|||    :::||||||:::|||||||||||    |||||
Db     1134 GAAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAA 1193

Qy      322 sValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu-------------- 337
              |||||||    ||||||      |||     |||    |||||||||:::
Db     1194 AGTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGA 1253

Qy      338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTy 353
                          |||||||||||||||         :::    |||||||||||||:::||
Db     1254 GCAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATA 1313

Qy      353 rGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGl 373
              |||||||||||||||||    |||:::||||||          :::      ||||||||
Db     1314 TGAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGA 1373

Qy      373 uLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Gl 392
              ||||    |||||||||||||||||||||||||||||||||:::|||     ||||||    ||
Db     1374 GAAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGA 1433
```

```
Qy    392 uValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGl 412
Db   1434 GGTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG-------- 1485

Qy    412 yGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLy 432
Db   1486 ----------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGA 1529

Qy    432 sLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSe 452
Db   1530 ATTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCT 1589

Qy    452 rArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGl 472
Db   1590 CAAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGA 1649

Qy    472 uLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGl 492
Db   1650 GAAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACA 1709

Qy    492 nGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLy 512
Db   1710 GGAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAA 1769

Qy    512 sLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGl 532
Db   1770 AAAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGG 1826

Qy    532 yGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSe 552
Db   1827 GGAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAG 1886

Qy    552 rSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGl 572
Db   1887 CTCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGA 1946

Qy    572 uAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGl 592
Db   1947 GGTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGA 2006

Qy    592 uGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgAr 612
Db   2007 AGGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACG 2066

Qy    612 gProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSe 632
Db   2067 GCCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTC 2126

Qy    632 rThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLy 652
Db   2127 CACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAA 2186

Qy    652 sProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVa 672
Db   2187 GCCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGT 2246

Qy    672 lGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProAr 692
Db   2247 GGGATCATCCAAGAAAAGAGCAAGGAGAGGGTCCTCTTCTGATGAGGAAGGGGGACCAAA 2306

Qy    692 gThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyTh 712
Db   2307 AGCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGAC 2366

Qy    712 rAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGl 732
Db   2367 AGACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGA 2426

Qy    732 uProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLe 752
Db   2427 GCAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTT 2486

Qy    752 uValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp-------- 769
Db   2487 AGTCACGCCAGGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCCTAGC 2546

Qy    770 ----SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVa 788
Db   2547 TGGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGT 2606

Qy    788 lSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGl 808
Db   2607 CTCAATCAAGAAGTTTATTCCTGGACTAAGGAAGAGAAAAGGCCAGATGGGAAACAAGAACA 2666

Qy    808 nAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAl 828
Db   2667 AGCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGC 2726

Qy    828 aValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Al 845
Db   2727 CGTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCCCAGCAAGC 2786

Qy    845 aGln-GlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865
Db   2787 CCAAAAAAAGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCA 2846

Qy    865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885
Db   2847 GCGAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCA 2906

Qy    885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905
Db   2907 TTATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAG 2966

Qy    905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---G 924
Db   2967 TAGAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAG 3026

Qy    924 luGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetV 944
Db   3027 AAGAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGG 3086

Qy    944 alThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaL 964
Db   3087 TCGTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCAT 3146

Qy    964 euArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerA 984
Db   3147 TGGGTGCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAG 3206

Qy    984 laValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluV 1004
Db   3207 CAGTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGG 3266

Qy   1004 alGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnA 1024
Db   3267 TGGAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGG 3326

Qy   1024 laValAlaAspLysValLysGluGluSerGlnValProAlaThr--------------- 1038
Db   3327 CAGTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATG 3386

Qy   1039 -----GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluG 1057
Db   3387 TGCTTCAGCCTGTGCAGAGAGCA-----------------GAGGCAGAAAGACCAGAAG 3428

Qy   1057 luAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValG 1077
Db   3429 AGCAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTC 3488

Qy   1077 lnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluS 1097
Db   3489 AGGAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAA 3548

Qy   1097 erLeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrC 1113
```

```
                 ||   |||       |||:::|||||      ::::::           |||   ||||
Db     3549 GCTTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTT 3608

Qy     1113 ysGlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValA 1127
            |||||                  :::|||   |||::::::  ||||||||||||:::
Db     3609 GTCAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCC 3668

Qy     1127 laProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaA 1147
               |||:::|||  ||||||  |||||||||||||:::||||||||||||:::||||
Db     3669 CCCCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCG 3728

Qy     1147 spSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysA 1167
            ||   |||            ||||||:::||||||   ::::::      :::
Db     3729 ACTTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATG 3788

Qy     1167 laThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysG 1187
                  |||:::       ||||||      ||||||   ||||||      |||||||||
Db     3789 AGGTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAG 3848

Qy     1187 luGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProG 1207
            ||   |||          |||::::::         |||||||||     |||:::
Db     3849 AGAGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAA 3905

Qy     1207 ly---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValL 1226
                    |||   ||||||   |||      :::|||::::::      |||   :::|
Db     3906 AGATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTC 3965

Qy     1226 euAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValL 1246
            ||:::||||||||     |||         |||   |||            ||||||   |
Db     3966 TGTCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCA 4019

Qy     1246 ys---------------GluGluGlnGluValPheValHisSerGly------ProAsnS 1259
            ||               |||   |||   :::   :::|||            :::
Db     4020 AAGACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTC 4079

Qy     1259 erGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnG 1279
              :::|||      :::|||            |||          |||   |||::::
Db     4080 GGGAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAA 4139

Qy     1279 luLysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyG 1294
            ::  ::::::   |||:::||||||                |||   |||   |
Db     4140 AGGATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAG 4199

Qy     1294 luMetGluThrAspValGluLysGluLysArgGlu---ThrLysProGluGlnValSerG 1313
            |||||         ||||||:::|||:::   :::       |||      |||:::|
Db     4200 AGATGGTAGTTCAAGTCGAAAGGGAGGAATACCAGTAGCAGAGCCCAACCCATGTGAATG 4259

Qy     1313 luGluGly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProV 1332
            |||||      |||   |||||||||          |||            :::   :
Db     4260 AAGAGAAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCC 4319

Qy     1332 alLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlyS 1352
            ::    |||:::::::::|||      :::         |||   :::   ||||||   ||||
Db     4320 TCCAGACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAA 4379

Qy     1352 er---ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerL 1371
            ||   |||            |||::::::|||   |||   :::::||||||||||||
Db     4380 GCCCTCCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCT 4439

Qy     1371 euAspThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrV 1389
              :::   :::   |||   ||||||   |||   |||:::   :::      ||||||
Db     4440 CTGAGGCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTG 4499

Qy     1389 alValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaG 1409
                  |||   ||||||||||||:::   |||         ||||||||||||:::   |
Db     4500 CCAACATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAG 4559

Qy     1409 luLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProL 1429
            |||||||||||                 :::            ||||||   ||||||
Db     4560 AGAAATCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCA 4619

Qy     1429 euGlyProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerT 1449
              ||||||:::   ||||||:::|||   |||:::||||||:::            :::

Db     4620 CAGGGCCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAG 4679

Qy     1449 hrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluG 1469
               |||      ||||||:::||||||   :::   |||   :::   :::|||:::|||
Db     4680 GCTTAAGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAG 4739

Qy     1469 luAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysV 1489
               |||:::      |||         :::            |||   |||   |||||||||:::
Db     4740 TCGATGAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGAGG 4793

Qy     1489 alLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsn-LysIleValLeu 1508
               |||:::   |||         |||||||||||||||:::||||||:::   |||:::|||
Db     4794 ATTTAGAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAAACTTGTCCAA 4853

Qy     1509 AsnValIleGlnThrAlaValAspGln-PheAlaArgThr---GluThrAlaProGluTh 1527
            |||:::||||||||||||||||||||   |||      ||||||   ||||||||   |||
Db     4854 AACATCATCCAGACAGCCGTTGACCAGTTTTGTACGTACAGAAGAAACAGCCACCGAAAT 4913

Qy     1527 rHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAs 1547
                        :::   |||||||||         |||   |||         |||   :::|||
Db     4914 GTTGACGTCTGAGTTACAGACACAA------GCTTCACGTGATAAAAGCTGACAGCCAGG 4967

Qy     1547 nArgCysTrpThrLysMetLysAspAlaLys--------------------------Me 1558
              |||   |||   :::   :::      |||                           ||
Db     4968 ACGC---TGGGACAGGAAACGGAGAAAGAAGGAGAGGACCCTCAGGCCTCTGCACAGGAT 5024

Qy     1558 tLysHisProValProGlnProArgGlu--------------------AspLeuGlnVa 1571
             |||||||    :::   ||||||:::                     |||      ::
Db     5025 GAAACACCAATTACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACAT 5084

Qy     1571 lLeuThrValLeuGluAlaTrpAlaGlnProArgLys 1583
             :|||         :::         :::||||:::|||
Db     5085 TCTGATATTTCCAAAGACATGAGTGAAGCCTCAGAAA 5121
```

RESULT 9
ABL23161

```
ID   ABL23161 standard; DNA; 8307 BP.
XX
AC   ABL23161;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 20956.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 20956; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
```

```
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 8307 BP; 2347 A; 2316 C; 2191 G; 1453 T; 0 other;

Alignment Scores:
Pred. No.:              6.76e-24      Length:         8307
Score:                  642.00        Matches:        444
Percent Similarity:     34.68%        Conservative:   250
Best Local Similarity:  22.19%        Mismatches:     735
Query Match:            7.95%         Indels:         572
DB:                     23            Gaps:           94

US-09-902-432-4 (1-1596) x ABL23161 (1-8307)

Qy        2 GlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThrPro 21
Db     1648 GGTGCCACTTCAGCTTCCGAAGAA---AGTGAGGAGCAGGATGAAGGCAAGTCTACTGAA 1704

Qy       22 SerGluLeuValLeuSerGlyHisGlyProAla------------------AlaGluAla 35
Db     1705 GCTCCAACCTCTGTCGATGACATTGAGCCGGCCAAGCCAACTGAATCCTCTGAGGAAGCT 1764

Qy       36 SerGlyAlaAlaGlyAspProAla---AspAlaAspProAlaThrLysLeuProGlnLys 54
Db     1765 AGCGGCGAGGGAGAAGATGTTGCTAAGGAAACCACACCTGCTGGAGAA------------ 1812

Qy       55 AsnGlyGlnLeuSerSerValAsnGlyValAlaGlu------GlnGlyAspValHisVal 72
Db     1813 ------------GCCAGCATTGCTGGTGAGGAGGAGATCGTTAAGGGAACCACACCAGCT 1860

Qy       73 GlnGluGluAsnGlnGluGlyGlnGluGluGluValValAspGluAspValGlyGlnArg 92
Db     1861 GGAGAACCTAGCAGTGAAGGC---GATGAGGAGATCGTTAAGGGAACAACTCCAGCAGAA 1917

Qy       93 GluSerGluAspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAla 112
Db     1918 GAATCCAGTAGTGAAAGTGAA------------GATGAGCTCACCAAGGTCACCACACCT 1965

Qy      113 ValGluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIleGluGlnIlePro 132
Db     1966 GCCGGAGAACCTAGCGTTGCTGGAGAAGAAGAGATTGCCAAG------GAAACCACACCT 2019

Qy      133 AlaSerGluAsnAsnVal---------GluGluMetValGln---------ProAlaGlu 146
Db     2020 GCCGGAGAACCTAGCATTGCTGGAGAAGAGGAGATCGTCAAGGTCACCACCCCCGCT--- 2076

Qy      147 SerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGlyPheLysPheThr 166
Db     2077 GGAGAATCTAGCATTGCTGGAGAAGAGGAGATC---------------GTCAAGGTCACC 2121

Qy      167 ValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAsp 186
Db     2122 ACCCCCGCTGGAGAATCTAGCAGTGAAGGCGAAGAGGAGATCATCAAGGTCACCACCCCC 2181

Qy      187 GluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGlu 206
Db     2182 GCTGGAGAA------------TCTAGCAGTGAAGGAGACGAAGAAATCGTTAAGGAAAGC 2229

Qy      207 ThrAlaValGlyGlu------SerAlaSerLysGluSerGluLeuLysGlnSerThrGlu 224
Db     2230 ACTCCTGCCGGAGAACCAATCAGTGAAGGCGAAGAGGATGTCATCAAGGCCACCACTTCA 2289

Qy      225 LysGlnGluGlyThrLeu---------LysGlnGluGlnSerSerThrGluIleProLeu 241
Db     2290 GCTCCCAAATCCGACATCGAAGGCGTCAAGGAACCTGAGACAGCCACAGAAGTACCAGCA 2349

Qy      242 GlnAlaGluSerAspGlnAlaAlaGlu------------GluGluAlaLysAspGlu--- 256
Db     2350 GAAGAAGTCGAGGACTTCGCCAAACCCACAACGCCAATCGCTGAGGCGGAGGAAGAACCC 2409

Qy      257 ------------------------------------GlyGluGluLys------------ 260
Db     2410 ATTGCCGGAACGCCTATTCCAACTGACGGTATCAGTGGCGAAGAGGAGATTGTTAAGGGC 2469

Qy      261 ---------------GlnGluLysGluProThrLysSerProGluSerProSerSerPro 275
Db     2470 ACCACTCCACAGACCCTCGAGGAGCAGCCTGAAATTAGTGAAGAGTCTACCGAGGTACCA 2529

Qy      276 Val---------------AsnSerGluThrThrSerSerPheLysLysPhePheThrHis 290
Db     2530 GTTGCTGAGGACGATCTTTCGAGCTCGACCAGTGCTTCAGCAATCGCCTCCTCGACCGAA 2589

Qy      291 GlyTrpAlaGlyTrpArgLysLysThrSerPheLys------------------------ 302
Db     2590 GGTGTCCAGGATGCTGCAAGTGAAACGACTTCCAGTGCACCTGCCAGAGCTGGAGACAAG 2649

Qy      303 ------------------------------LysSerLysGluAspAsp------------ 308
Db     2650 GACGAAGCAGCTACCACCGTGCCTACTGCCCAAGACAAAGATGATGAAGAGGTTGAACAG 2709

Qy      309 ---------------------------------------------------LeuGluThr 311
Db     2710 GATGCCACAGACTTGCCAGTCGAGGATGTTGTCCAAAGCACAACTGCAAAGACCACAACC 2769

Qy      312 AlaGluLysArgLysGluGln---------GluAlaGlu--------------------- 321
Db     2770 ACCGAGCAGCCCAAGGAGGAGTCCTCTACTGAGGCTGAGGACGCTGAAATCGAAGTTACC 2829

Qy      322 ---------LysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGln 338
Db     2830 ACTAGTAGTCCCGCCGATAAACAAGAGGTGCCAGAAGCTGAACCAGCAGACAAGGACCAT 2889

Qy      339 GluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeu 358
Db     2890 AAGGATGAGGAAGATGTACAGACCGCC------------------------ACTGATTTG 2925

Qy      359 ProLeuGluAspGlnValGly------------AspLeuGluAla--------------- 369
Db     2926 CCAATCAAATCTGATATTGGGCCACCTGTTGTCGATACCGAAGCTACTACGGGTCAGCCT 2985

Qy      370 ---SerSerGluGluLysCysAla---------------------------ProLeuAla 379
Db     2986 GAGACCAGTGATGAGACTGCTACCGATAAGCCACCATCTGTTTATCTGCCTCCTGTGTCC 3045

Qy      380 ThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisVal 399
Db     3046 CAGGAAGTTCCGTCCAGCACTGCTAAAGTTGACAATCGCAACGACTTCGAAACGGAGAAA 3105

Qy      400 SerThrValGluLysThrGluGluGluGlnGly--------------------------- 410
Db     3106 CCAACCCTGCCACCAAGTGGCGAGGATCAGAGCTCCGAGCCCCTGCCCGCCATGGATCTG 3165

Qy      411 ---GlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuPro 429
Db     3166 CCCGCTGGTATTCCTGGCGAAGGTGACTGTCTGGTGGAGGGC------------------ 3207

Qy      430 ProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGlu------ 447
Db     3208 ---AAGACCTATGCCAACAACACCATCGTTCCCGCTACCGCA---CCCTGCGACGTGTCC 3261

Qy      448 ---------------------------------------------GluLeuMetLysSer 452
Db     3262 TGCAAATGTATCAGCAGCCTCGTGGCCTGCCAACAGATGGAGTGCAAGTTGCCCGAGAAC 3321

Qy      453 ArgGluMetCysValSerGlyGlyAsp------------------HisThrGlnLeuThr 466
Db     3322 TTGGAGAAATGTACCGTGGCTGCCGATCTTCTGGATGGTTGCTGTCCCACTTACATTTGC 3381

Qy      467 AspLeuSerProGlu---------------GluLysThrLeuProLysHisProGluGly 481
```

```
Db   3382 GATGAGAGCACTGAAAGTGCTGAAAAGGATGAGGAATCTACCGCCAAGCCCGACAATAAG 3441
Qy    482 IleValSerGluValGluMetLeuSerSerGlnGluArgIleLys--------------- 496
Db   3442 ATTGACGAGGATGTTTCAGAGATCAGCACTGAAGAGATTCCCAAGGATGTCATCATGCCT 3501
Qy    497 ------ValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeu 514
Db   3502 ACTGGCATCACTGAGCAGCCTCTATCTCATGTC--------------------------- 3534
Qy    515 SerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyr 534
Db   3535 ---------------------------------------------AAGCCCGATGAAGAG 3549
Qy    535 GlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAla 554
Db   3550 ATCCAACCAGTGACCTCGGTTCCTGCACAGTTTGATGAA---------TCTACGACAGCT 3600
Qy    555 SerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaPro 574
Db   3601 AAGGTTGATAAGAAACCAATTGATGAATCCGCTGAGGATAAGAAACCAATT--------- 3651
Qy    575 GlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIle 594
Db   3652 ------GGCGAATCCGAGGAGGATTCGAAACCAATTGATGAATCTGAGGAGGAC------ 3699
Qy    595 ThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArgProSer 614
Db   3700 ------------------AAGAAACCGGTTGAAGAATCTGCAGAAGATAAGAAACCCGTT 3741
Qy    615 GluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAsp 634
Db   3742 GAA---GATTCCGAAGAGAAGGAGAAGCCATTGCCGACTGTCATCCCAGCGAGTGAGATT 3798
Qy    635 SerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLysPro--- 653
Db   3799 GAAAAAGAATCCAAGCCCGAGGACGAAAAGAAGACTGAGGCTGACTTTGCGGCGCCCACT 3858
Qy    654 GluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGly 673
Db   3859 GAACAGCCAGAGGCTACAACTCCCGCCCAGATTGCCGACACTGCTGAGAAGGAGGTTGAT 3918
Qy    674 SerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThr 693
Db   3919 GATAAGCTGGCAACCACATCTGCTCCCGTTTCTGGCGAAGATGAGCTCAAGCCT------ 3972
Qy    694 LeuGlyGlyAspSerHisArgAlaGluGluAlaSer---LysAspLysGlu------Ala 710
Db   3973 ---GCCGATGAAAAGAAACGCACTGAGACAGCGCAGATCCCAGATGCTGAAATCCCTGCT 4029
Qy    711 GlyThrAspAlaValProAlaSerThrGln------------------------------ 720
Db   4030 AGCACCGATGAGCCCGAGTCCTCGACTGAGCTGCCTACTGTTGATCTCGACAAGAAACCA 4089
Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGlu---------------------Pro 733
Db   4090 GAGGAGGATTCGACTAAAGGCACCGAGGCGCCTGAGTCTGATAAGGTTCCTGAAGTTCCG 4149
Qy    734 AlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuVal 753
Db   4150 ACCAGTGCTTCTACTGAAAATGAG---ATTGAGGAGTCTGACAAATTCACGACAGTCGCT 4206
Qy    754 ThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGlu 773
Db   4207 CCACCAAAAAATTTCTGCTTCTGACGAAACAGAGCCCACTGCTGAGGAAGATCTCGTC--- 4263
Qy    774 GlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLys--- 792
Db   4264 ------CCTGCAACCTTTGAGCCCATCGAAAGCGAATTTGAAGTTAGCACCAAGAAGCCA 4317
Qy    793 ---------------------------PheIleProGlyArgArgLysLysArg------ 801

Db   4318 GCCGTGCAAGGACCACCACTGCCCACTCTTGCTCCGGCTCAGCCAGAGAAGAAACCAGTC 4377
Qy    802 ---------------------AlaAsp---GlyLysGlnGluGlnAlaThrValGluAsp 813
Db   4378 GATGCAGAGACTTCCACTGAGGCTGATATCAGCACCGAACCAAGCGCTGAAGTTGAAAAG 4437
Qy    814 SerGlyProValGluIleAsnGluAspAspProAsnVal--------------------- 826
Db   4438 GAAGCTTCTGGTGAAACATCCGAGTCGGACAATGAAATCGATGCTGGTGCTTCTAGTACT 4497
Qy    827 ProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLys---MetGluAla 845
Db   4498 CCTGTTCCTGTTTCTGCCGATGAGGACAAGACTCCAAGCACTGAGAAGACTGTCGAGGCC 4557
Qy    846 ---------------------------------------GlnGlyAsnThrGluLeuPro 852
Db   4558 GACGACAAGTTCACCACTGTTGCGCCACTTGCTGGTGATGAAGAGGAGTCCAACTTGCCC 4617
Qy    853 GlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrVal 872
Db   4618 AAGCTACCACAAGATATCTTT------GAAGAGGAAGCGCCTGTTGCGGTTACCACCGCA 4671
Qy    873 SerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArg--------- 889
Db   4672 GCTCCATCAAAGGACGACGGTGAGCAGAAGCCAGTCGAAGTTGAAGAGAAACCGATCGAA 4731
Qy    889 ------------------------------------------------------------ 889
Db   4732 GATGGACAAAAGCCTATTGAGGACGAAACCTCTACACCTACTTCATCTGAAAATGAGATT 4791
Qy    890 ---------------SerProSerTrpIleSerAlaSerValThrGluProLeuGluHis 904
Db   4792 GAGCCTGAATCTGACCGTGCAACTACTATTGCTCCTTCTAAGGAAGAACCTTCCGAACCA 4851
Qy    905 ThrAlaGly---------------GluAlaMetProProValGluGluValThrGluLys 919
Db   4852 TCCACTGGTGCTCCTACTAAGGATGAACCTGCCGAACCATCCACTGATGCTCCTGAGTCT 4911
Qy    920 AspIleIleAlaGluGluThrPro---------ValLeuThrGlnThrLeuProGluGly 936
Db   4912 GAT---GAGAGCAAGGAAACGCCTGAGTCCGAAGTGCCCACAACTGTTGCACCTGCTGGT 4968
Qy    937 LysAspAlaHisAspAspMetValThrSerGluValAspPheThrSerGluAlaValThr 956
Db   4969 GAGAAGATTCCAACAAGCAGTATCACTCCCGATGAGGAACCAACTGCCACTTCTGCTCCT 5028
Qy    957 AlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGlu--------------- 971
Db   5029 GTTGCCAAGCCTGATGAAGATGTAGAGAAAGAAACCAGCACCGAAATTCCCACTGACGCT 5088
Qy    972 ---AlaSerGlyAlaGluGlu---------ThrThrAspMetValSerAla--------- 984
Db   5089 CCTGCATCGTCCGAAGAGGATGAGAATTCCTCTACTGATCAAATTCCATCTGAGGTACCA 5148
Qy    985 ---------------------------------------------ValSerGlnLeuThr 989
Db   5149 GAAAAGAAACCTGAAACTCCTGCACAGACGCCAGAAGAAGGAGATATTGTTGGCGCAACT 5208
Qy    990 AspSerProAspThrThrGluGluAlaThrProValGlnGluVal--------------- 1004
Db   5209 GCAGCTCCAACGACCTCTGATGAAGTGCCCCCGGTCCAGAGGCTGCCTGAGGAAGTCCTT 5268
Qy   1005 ------------------------GluSerGlyValLeuAspThrGluGlu--------- 1013
Db   5269 GCTGAGATTCCGCAGCCCTCGACTGAAACTGGAATTAAACAGCAAGACGAAACAACTGCT 5328
Qy   1014 ---------------------------------------------GluGluArgGlnThr 1018
Db   5329 GCTCCAAGTATTGACCGCAAGGAACCCTATGTTACTGAAATCGATGAAGAAGCCACCACA 5388
Qy   1019 GlnAlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThr 1038
Db   5389 GTGGCTCCTATCTCCGAGAAGGATGAGAAGCCAACTGAGGAA------------------ 5430
```

```
Qy   1039 GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAsp 1058
                                     |||   :::|||:::            ||||||
Db   5431 ---------------------GAGAAACCAGTGGAACAGAAGCCAACTGGAGAAGAGCCA 5469

Qy   1059 SerGluValLeuAlaSerGluLys------GluLysAspValMetProLysGlyProVal 1076
          ||||||              ||||||       |||:::||||||      :::|||||||||
Db   5470 TCCGAGGAAGAAGAGAAGGAGAAACCCATTGAACAAGATGTATCTACTGAGGGACCAGTA 5529

Qy   1077 Gln---GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrPro 1095
                ||||||                   :::      ||||||      :::   :::
Db   5530 TCTACCGAAGCTTCTGAAGCTGGATCTACCGAATCATCTGAGGAAGTTAAGCCCTCGACT 5589

Qy   1096 GluSerLeuGluValProGluValThrAlaAspValAspHisValAlaThrCysGlnVal 1115
          |||      ||||||   |||          |||             :::|||   |||
Db   5590 GAGGGA---GAAGTTGCCGAAAAGCCTGAGGACAAGCAGCCTTCTTCGACTGCCCAGGCT 5646

Qy   1116 IleLysLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeu--- 1134
                                                ||||||   |||      |||
Db   5647 ---------------------CCAGTTGAGACAATTCCCGAAATCTCCACTGAACTGCCA 5685

Qy   1135 ---ThrAspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGly 1153
                |||   :::         |||         |||||||||
Db   5686 GCTCAAGATGGCGACAAGCCCACAAGCGAGGCTCCTGTTGACTCCGATGAAGATACCTCC 5745

Qy   1154 ThrGlnGlnAspGluThrIleAspSer-----------------GlnAspSerLysAla 1167
                ||||||   |||   |||                       :::      :::
Db   5746 GCTCCGTCCGACGAGAAGATTCCCTCCGTTTCTGGCGAGGAAGTCGAGGGCCCAGAAGTC 5805

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
          |||   |||         |||:::         |||:::|||      |||      :::
Db   5806 ACAACTGCCTCACCTCAAGCT---GCCGAAGAGGACGAACTTAAGACACCCGCGGAGAGT 5862

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
          |||||||||:::         :::   ||||||      |||            |||
Db   5863 GAACCATCTTCC---ACTGACAAGGTACCAGAAACTGAATACCAGAAGCCCGAG------ 5913

Qy   1208 ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAla 1227
                         :::   |||:::   :::   |||      :::|||      ::::::
Db   5914 ------------GACGAGACCAAGGCTGACGAGACCCCTGAGTCTGTTACCCAAGTGTCT 5961

Qy   1228 LysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLys--------- 1244
                                     |||      |||   ||||||
Db   5962 GATGTGGCAACATCCACTTCGGCCCCAGTTGCTGGTGGTGACATTGAAAAGGACGAACAA 6021

Qy   1245 ---------------ValLysGluGluGlnGluValPheValHisSerGlyProAsnSer 1259
                         :::||||||:::|||:::                  |||
Db   6022 GCCACCACTGCCTCGCCTGAAGAGGAGGAAGAGATC-----------AAGCCAACTATC 6069

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
                ||||||::::::            |||                           |||
Db   6070 GCTCCAGCTGCTGAAATTCCTCAACCATCC--------------------------GAA 6102

Qy   1280 LysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspVal 1299
          ||||||      :::   |||         :::|||   |||         |||
Db   6103 AAGGAGCCCGTGGATGAGCAA------GAAGTGGAATCCGGAACCAAGGCCACTCCAGCA 6156

Qy   1300 GluLysGluLysArg---------GluThrLysProGluGlnValSerGluGluGlyGlu 1316
          |||   :::   :::            |||      |||                  |||
Db   6157 GAGTCTGACGGCCAGCCAATTGATGAAATTGCCCCTGCAACATCTGGACCCATCGATGAG 6216

Qy   1317 GlnGluThrAlaAlaProGluHisGluGly------------------------------ 1326
                |||||||||||      |||
Db   6217 GCTTCCACAGCCGCTCCCACAAAGGAGGAATCCACAACCGTTGCTAGCGCAGCTTCGCCT 6276

Qy   1327 ------------------------------ThrTyrGlyLysProValLeu--------- 1333
                                        |||      :::||||||
Db   6277 GCTGTTCACGATGATGAAATTAAGGATGTCACCACTACGCAACCAGTTGCCGATGAGAAG 6336

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerPro 1353
             :::      |||      |||         ::::::            ||||||
Db   6337 GAAGTGGCTGCTCCTCAGGACGAAACGAAGACATCCATTGATGTTTCCACCGATTCACCC 6396

Qy   1354 SerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThr 1373
          :::      |||   :::|||            |||      |||            |||
Db   6397 ACTGCCCAAGATGATGAGAAGCAAGACAAAACCGAGGCTCCAGTG------GCACCCACC 6450

Qy   1374 ThrVal------------------------------------ThrGlnThrAlaGluAla 1381
          ||||||                                    |||   |||   |||
Db   6451 ACTGTTTCATCGCCGACTGCTGATTCTGCTGCTGATTCTAGTACTCCCACAGTGGAGGTA 6510

Qy   1382 ValGluLysVal---IleGluThrVal------------ValIleSerGluThrGlyGlu 1396
                |||   |||:::|||                  ::::::|||:::|||
Db   6511 CCATCTCCTGTTGAAATCGACACGAAGCCAATGGATGATATCATGTCCCAGACTATTGCT 6570

Qy   1397 SerProGluCysValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGly 1416
                         ||||||               |||
Db   6571 CCTCATACCGCCGATGGTGCTGCGTCCACGTCCACTGAAGATGAAGACCAGGCTCCCGTT 6630

Qy   1417 HisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyProGluSerGln------ 1434
                :::   |||   ||||||   |||   |||:::   |||         |||
Db   6631 ACTGTGTCTCCTCAGGATGCTGAAAAGACT---CCAGTTTCCCCGGCTCCCCAGGATTCT 6687

Qy   1435 ---------------------------------AlaGluSerIleProIleIleValThr 1443
                                           |||:::   ||||||      |||
Db   6688 GATAAGACTCCGAGCTCAGAGGCTCCCCAGGATGCCGACGAAATCCCAGCTACAGCGACA 6747

Qy   1444 ProAlaProGlu------------SerThrLeuHisProAspLeuGlnGlyGluIleSer 1459
          |||   :::               :::|||:::   |||                :::
Db   6748 CCTCTTGACGACAACAAGATCCCAGCCACTGTGGCTCCTCAGACTGATGATGGGGTTCCA 6807

Qy   1460 AlaSerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyPro---AspAla 1478
          |||:::               |||::::::   :::      |||   |||   |||
Db   6808 GCTACCGCAGCTCCTCTTGATGAAGACAAGATCCAAACTACTGCGGCTCCTCTTGATGAG 6867

Qy   1479 AspGlyLysGluSerThrAlaIle------------GluLysVal---LeuLysAlaGlu 1493
          :::         |||||||||               ||||||:::
Db   6868 GAGAAGATTCCATCCACAGCAGCTCCTCTTGATGATGAAAAGATTCCTGCTCCAGTTAGC 6927

Qy   1494 ProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnValIleGlnThr 1513
          |||   :::   ::::::|||      |||   |||
Db   6928 CCCGTTGTTTTCGATGTGGAGCCAAGCAGCGAGAAG---------------------CCT 6966

Qy   1514 AlaValAspGlnPheAlaArgThrGlu---ThrAlaProGluThrHisAlaTyrAspSer 1532
          ||||||   ::::::         |||   |||   |||      |||      ::::::
Db   6967 GCTGTATCTGAGTATGATGGTGAAGAGAGCACCGAACCACCTGTCCATGATGTGGAGACC 7026

Qy   1533 GlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCysTrpThrLys 1552
             |||   |||
Db   7027 AGCACTGATGAGCCTACT------------------------------------------ 7044

Qy   1553 MetLysAspAlaLysMetLysHisProValProGlnProArgGluAspLeuGlnValLeu 1572
                |||||||||:::|||   |||         |||            :::
Db   7045 ---TCTGATGCCAAACTGAAGCCACCTACTTCTGCACCTGCTACGCCATCGGAA---TCT 7098

Qy   1573 ThrValLeuGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAla 1592
                   ||||||      |||         |||   |||:::   :::
Db   7099 CCAGCCACGGAGGCGGAGATTGTGCCAGAAACTGCAGCCCCAGAGCTTGAAAAGGAGGTG 7158

Qy   1593 Pro 1593
          |||
Db   7159 CCA 7161
```

```
RESULT 10
ABL23160
ID   ABL23160 standard; DNA; 11597 BP.
XX
AC   ABL23160;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 20953.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
```

```
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 20953; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 11597 BP; 3279 A; 3082 C; 2895 G; 2341 T; 0 other;

Alignment Scores:
Pred. No.:              1.29e-23      Length:         11597
Score:                  639.50        Matches:        389
Percent Similarity:     34.82%        Conservative:   227
Best Local Similarity:  21.99%        Mismatches:     707
Query Match:            7.92%         Indels:         447
DB:                     23            Gaps:           69

US-09-902-432-4 (1-1596) x ABL23160 (1-11597)

Qy      3 AlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThrProSer 22
             |||   ::::::|||   ::::::      |||:::   :::          :::
Db   4870 GCAAAGACCACAACCACCGAGCAGCCCAAGGAGGAGTCCTCTACTGAGGCTGAGGACGCT 4929

Qy     23 GluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspPro 42
             |||:::   ::::::            ||||||   :::             :::|||
Db   4930 GAAATCGAAGTTACCACTAGTAGTCCCGCCGATAAACAAGAGGTGCCAGAAGCTGAACCA 4989

Qy     43 AlaAspAlaAsp-----------------------ProAlaThrLysLeuProGlnLys 54
             ||||||   |||                       ||||||   ||||||   |||
Db   4990 GCAGACAAGGACCATAAGGATGAGGAAGATGTACAGACCGCCACTGATTTGCCAATCAAA 5049

Qy     55 AsnGly------------------GlnLeuSerSerValAsnGlyValAlaGluGlnGly 68
             :::                             :::      |||      |||
Db   5050 TCTGATATTGGGCCACCTGTTGTCGATACCGAAGCTACTACGGGTCAGCCTGAGACCAGT 5109

Qy     69 Asp--------------------------ValHisValGln---------GluGluAsn 76
             |||                          |||::::::         |||::::::
Db   5110 GATGAGACTGCTACCGATAAGCCACCATCTGTTTATCTGCCTCCTGGTGGAGAGAAGGAT 5169

Qy     77 GlnGluGlyGlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAsp 96
                |||   ::::::|||            |||         |||   |||
Db   5170 ACTGAGAAAGAGGACGAAACTACTGCTACAGAGCCATCTGTCCAAGAGGAACCATCTAAG 5229

Qy     97 ValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIle 116
                :::         |||   :::         :::|||||||||      |||
Db   5230 CCTGTGGATATCTCTCCAGTGTCCCAGGAAGTTCCGTCCAGCACTGCTAAAGTTGACAAT 5289

Qy    117 ThrLysAsp---------------------------------GlyGlnGluGluThrSer 125
                |||                                 |||::::::::::::|||
Db   5290 CGCAACGACTTCGAAACGGAGAAACCAACCCTGCCACCAAGTGGCGAGGATCAGAGCTCC 5349

Qy    126 GluIleIleGlu------GlnIleProAlaSerGluAsnAsnValGluGluMetValGln 143
          |||   :::          :::||||||                  :::   :::
Db   5350 GAGCCCCTGCCCGCCATGGATCTGCCCGCTGGTATTCCTGGCGAAGGTGACTGTCTGGTG 5409

Qy    144 ProAlaGluSerGlnAlaAsn-----------------------------AspValGly 153
                ::::::   ||||||                               ||||||
Db   5410 GAGGGCAAGACCTATGCCAACAACACCATCGTTCCCGCTACCGCACCCTGCGACGTGTCC 5469

Qy    154 PheLysLysValPheLysPheValGlyPhe------------------------------ 163
             |||   :::         |||
Db   5470 TGCAAATGTATCAGCAGCCTCGTGGCCTGCCAACAGATGGAGTGCAAGTTGCCCGAGAAC 5529

Qy    164 ------LysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeu 181
                |||   ||||||      |||         :::               :::
Db   5530 TTGGAGAAATGTACCGTGGCTGCCGATCTTCTGGATGGTTGCTGTCCCACTTACATTTGC 5589

Qy    182 ThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGln 201
                   :::      |||      |||   |||               |||::::::
Db   5590 GATGAGAGCACTGAAAGTGCTGAAAAGGATGAGGAATCTACCGCCAAGCCCGACAATAAG 5649

Qy    202 GluProSerValGluThrAlaValGlyGluSerAlaSerLysGlu--------------- 216
                ::::::         |||   |||   :::::::::|||
Db   5650 ---------ATTGACGAGGATGTTTCAGAGATCAGCACTGAAGAGATTCCCAAGGATGTC 5700

Qy    217 ------------------SerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeu 230
                            :::|||         |||      |||   :::
Db   5701 ATCATGCCTACTGGCATCACTGAGCAGCCTCTATCTCATGTCAAGCCCGAT--------- 5751

Qy    231 LysGlnGluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGlu 250
          ::::::   |||      |||   :::|||   |||   :::            |||:::
Db   5752 GAAGAGATCCAACCAGTGACCTCGGTTCCTGCACAGTTTGATGAATCTACGACAGCTAAG 5811

Qy    251 GluGluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGlu 270
             :::   |||      |||      ::::::|||:::|||               |||
Db   5812 GTTGATAAGAAACCAATTGATGAATCCGCTGAGGATAAGAAACCAATTGGCGAATCCGAG 5871

Qy    271 SerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHis 290
                |||   |||::::::
Db   5872 GAGGATTCGAAACCAATTGAT--------------------------------------- 5892

Qy    291 GlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGlu 310
                                                      |||:::|||   :::
Db   5893 ---------------------------------------GAATCTGAGGAGGACAAGAAA 5913

Qy    311 ThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGluLys 330
                |||:::   |||   :::            |||   ||||||||||||
Db   5914 CCGGTTGAAGAATCTGCAGAAGATAAGAAACCCGTTGAAGATTCCGAAGAGAAGGAGAAG 5973

Qy    331 ------------ThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGln 346
                      ||||||||||||   ::::::   ::::::         |||:::
Db   5974 CCATTGCCGACTGTCATCCCAGCGAGTGAGATTGAAAAAGAATCCAAGCCCGAGGACGAA 6033

Qy    347 AlaArgLeuSerAlaAspTyr------------GluLysValGluLeuProLeuGluAsp 362
             :::      ||||||:::              |||:::   |||
Db   6034 AAGAAGACTGAGGCTGACTTTGCGGCGCCCACTGAACAGCCAGAGGCTACAACTCCCGCC 6093

Qy    363 GlnValGlyAsp---LeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGlu 381
          |||:::   |||      |||         ::::::|||   |||         :::
Db   6094 CAGATTGCCGACACTGCTGAGAAGGAGGTTGATGATAAGCTGGCAACCACATCTGCTCCC 6153

Qy    382 ValPheAspGlu------------------------------LysMetGluAlaHisGln 391
          |||   |||                                 :::   |||      |||
Db   6154 GTTTCTGGCGAAGATGAGCTCAAGCCTGCCGATGAAAAGAAACGCACTGAGACAGCGCAG 6213

Qy    392 GluValValAlaGluValHisValSerThrValGlu------------------------ 403
                ||||||:::         ||||||   |||
Db   6214 ATCCCAGATGCTGAAATCCCTGCTAGCACCGATGAGCCCGAGTCCTCGACTGAGCTGCCT 6273
```

```
Qy     404 ------------------LysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGly 417
                             |||   ||||||:::       |||   ||||||
Db    6274 ACTGTTGATCTCGACAAGAAACCAGAGGAGGATTCGACTAAAGGCACCGAGGCG------ 6327

Qy     418 GlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGln 437
                                                ||||||        |||
Db    6328 ------------------------------------CCTGAGTCTGATAAGGTTCCT--- 6348

Qy     438 GluValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCys--- 456
           ||||||||       |||             |||:::   :::|||   :::
Db    6349 GAAGTTCCGACCAGTGCTTCTACTGAAAATGAGATTGAGGAGTCTGACAAATTCACGACA 6408

Qy     457 ---------------ValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGlu 471
                          :::|||      |||   |||:::   |||          |||
Db    6409 GTCGCTCCACCAAAAATTTCTGCTTCTGACGAAACAGAGCCCACT---------GCTGAG 6459

Qy     472 GluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSer 491
           |||      :::|||         |||   |||   ||||||   |||:::      |||
Db    6460 GAAGATCTCGTCCCTGCAACCTTTGAGCCCATCGAAAGCGAATTTGAAGTT------AGC 6513

Qy     492 GlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeu 511
              ::::::      |||||||||   ||||||      |||      :::
Db    6514 ACCAAGAAGCCAGCCGTGCAAGGACCACCACTGCCCACTCTTGCTCCGGCTCAGCCAGAG 6573

Qy     512 LysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluPro 531
           ||||||                                          |||   |||
Db    6574 AAGAAA------------------------------------CCAGTCGATGCAGAGACT 6597

Qy     532 GlyGluTyrGlnHisIleHisThrGlu---SerProGluSerAlaAspGluGlnLysGly 550
                       |||   ||||||   |||   |||         |||      |||
Db    6598 TCCACTGAGGCTGATATCAGCACCGAACCAAGCGCTGAAGTTGAAAAGGAAGCTTCTGGT 6657

Qy     551 GluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyPro 570
           |||:::|||   |||      |||                     ::::::   |||
Db    6658 GAAACATCCGAGTCGGACAATGAA---------------------ATCGATGCTGGTGCT 6696

Qy     571 LeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLys 590
                 |||            |||:::|||         |||      ||||||
Db    6697 TCTAGTACTCCTGTTCCTGTTTCTGCCGATGAGGACAAGACTCCAAGCACTGAGAAGACT 6756

Qy     591 ArgGluGly------------------IleThrProTrpAlaSerPheLysLysMetVal 604
             |||                      :::   |||   |||      ::::::
Db    6757 GTCGAGGCCGACGACAAGTTCACCACTGTTGCGCCACTTGCTGGTGATGAAGAG------ 6810

Qy     605 ThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLys 624
                                 :::   |||   :::      :::||||||
Db    6811 ---GAGTCCAACTTGCCCAAGCTACCACAAGATATCTTTGAAGAGGAAGCGCCTGTTGCG 6867

Qy     625 ValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluVal 644
           |||   :::|||      |||      |||               ::::::   |||
Db    6868 GTTACCACCGCAGCTCCATCAAAGGACGACGGTGAGCAGAAGCCAGTCGAAGTTGAAGAG 6927

Qy     645 LysThrValGlyGluGluGlnLysProGluGluProLysArgArgValAspThrSerVal 664
           |||   :::   :::   |||||||||   |||   :::            ||||||
Db    6928 AAACCGATCGAAGATGGACAAAAGCCTATTGAGGACGAAACCTCTACACCTACTTCATCT 6987

Qy     665 SerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSer 684
              |||               :::            ||||||         :::
Db    6988 GAAAATGAG------------ATTGAGCCTGAATCTGACCGTGCAACTACTATTGCTCCT 7035

Qy     685 SerAspAspGluGlyGlyProArgThrLeuGlyGly------------------------ 696
           |||   :::|||               :::   |||
Db    7036 TCTAAGGAAGAACCTTCCGAACCATCCACTGGTGCTCCTACTAAGGATGAACCTGCCGAA 7095

Qy     697 AspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValPro 716
             |||      |||   |||:::   :::      ||||||               ||||||
Db    7096 CCATCCACTGATGCTCCTGAGTCTGATGAGAGCAAGGAAACGCCTGAGTCCGAAGTGCCC 7155

Qy     717 AlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySer 736
           :::                  :::         :::||||||      |||
Db    7156 ACAACTGTTGCACCTGCTGGTGAGAAGATTCCAACAAGCAGTATCACTCCCGATGAGGAA 7215

Qy     737 ProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArg 756
           |||                                                |||
Db    7216 CCA------------------------------------------------ACTGCCACT 7227

Qy     757 LysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSer 776
                       :::|||   :::|||         |||      :::      ::::::
Db    7228 TCTGCTCCTGTTGCCAAGCCTGATGAAGATGTAGAGAAAGAAACCAGCACCGAAATTCCC 7287

Qy     777 ThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIleProGly 796
           |||:::            |||   ||||||
Db    7288 ACTGACGCTCCTGCATCGTCCGAAGAG--------------------------------- 7314

Qy     797 ArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyPro 816
                                       ::::::   :::         |||      |||
Db    7315 ------------------------------GATGAGAATTCCTCTACTGATCAAATTCCA 7344

Qy     817 ValGluIleAsnGluAspAspProAsnValProAlaValValProLeuSerGluTyrAsn 836
              |||:::   |||         |||      ||||||      |||
Db    7345 TCTGAGGTACCAGAAAAGAAACCTGAAACTCCTGCACAGACGCCA--------------- 7389

Qy     837 AlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGly 856
                    |||:::   :::         |||   |||         |||
Db    7390 ------------GAAGAAGGAGATATTGTTGGCGCAACTGCAGCTCCAACGACCTCTGAT 7437

Qy     857 AlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrVal------------ 872
              |||         :::   |||   :::   ::::::      :::
Db    7438 GAAGTGCCCCCGGTCCAGAGGCTGCCTGAGGAAGTCCTTGCTGAGATTCCGCAGCCCTCG 7497

Qy     873 ---------SerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArg 889
                    :::                     |||   |||   |||      |||::::::   :::
Db    7498 ACTGAAACTGGAATTAAACAGCAAGACGAAACAACTGCTGCTCCAAGTATTGACCGCAAG 7557

Qy     890 SerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAla 909
              |||                     |||||||||      |||
Db    7558 GAACCC---------------TATGTTACTGAAATCGATGAAGAAGCCACC------ACA 7596

Qy     910 MetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProValLeu 929
           :::      |||:::   |||      ||||||               ||||||   ||||||
Db    7597 GTGGCTCCTATCTCCGAGAAGGATGAGAAGCCA---ACTGAGGAAGAGAAACCAGTG--- 7650

Qy     930 ThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSer--------- 946
              :::         |||   |||::::::         ::::::
Db    7651 ---GAACAGAAGCCAACTGGAGAAGAGCCATCCGAGGAAGAAGAGAAGGAGAAACCCATT 7707

Qy     947 GluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThr 966
           |||   |||      ::::::|||         :::||||||   |||||||||      |||
Db    7708 GAACAAGATGTATCTACTGAGGGACCAGTATCTACCGAAGCTTCTGAAGCTGGATCTACC 7767

Qy     967 GluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSer 986
           |||
Db    7768 GAA--------------------------------------------------------- 7770

Qy     987 GlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGluSer 1006
                                 ::::::||||||      |||      |||
Db    7771 --------------------TCATCTGAGGAAGTTAAGCCCTCGACTGAG---------- 7800

Qy    1007 GlyValLeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAla 1026
           |||   :::      :::   |||:::::::|||      :::      ||||||
Db    7801 GGAGAAGTTGCCGAAAAGCCTGAGGACAAGCAGCCTTCTTCGACTGCCCAGGCTCCAGTT 7860

Qy    1027 AspLysValLysGlu---GluSerGlnValProAlaThrGlnThrValGlnArgThrGly 1045
           :::   :::   |||         :::::::::||||||            :::   |||
Db    7861 GAGACAATTCCCGAAATCTCCACTGAACTGCCAGCTCAAGATGGCGACAAGCCCACAAGC 7920

Qy    1046 SerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGlu 1065
                 ::::::                  :::||||||:::            :::   |||
Db    7921 GAGGCTCCTGTTGACTCC------------GATGAAGATACCTCCGCTCCGTCCGACGAG 7968

Qy    1066 LysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAla 1085
           |||         |||                           :::|||   |||
Db    7969 AAGATTCCCTCCGTT------------------------TCTGGCGAGGAA------GTC 7998

Qy    1086 GlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluVal------ 1103
```

```
                 :::|||  |||          |||:::|||:::::  |||      |||:::
Db    7999 GAGGGCCCAGAAGTCACAACTGCCTCACCTCAAGCTGCCGAAGAGGACGAACTTAAGACA 8058

Qy    1104 ---------------ThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeu 1118
                            :::         |||   |||       :::      |||
Db    8059 CCCGCGGAGAGTGAACCATCTTCCACTGACAAGGTACCAGAAACTGAATACCAGAAGCCC 8118

Qy    1119 GlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGlu 1138
           :::         :::               ||||||||         ::::::|||
Db    8119 GAGGACGAGACCAAGGCTGACGAGACCCCTGAGTCTGTTACCCAAGTGTCTGATGTGGCA 8178

Qy    1139 ThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGlu 1158
           |||:::   |||   |||:::|||       |||
Db    8179 ACATCCACTTCGGCCCCAGTTGCTGGTGGTGACATT------------------------ 8214

Qy    1159 ThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGlu 1178
                          :::|||  :::||||||  |||         |||       |||
Db    8215 ---------GAAAAGGACGAACAAGCCACCACTGCC---------TCGCCTGAAGAGGAG 8256

Qy    1179 GluGlu---------------AlaAlaThrAlaGlnLysGluGluProSerThrLeuPro 1193
           ||||||               |||      |||:::      :::||||||
Db    8257 GAAGAGATCAAGCCAACTATCGCTCCAGCTGCTGAAATTCCTCAACCATCC--------- 8307

Qy    1194 AsnAsnValProAlaGlnGluGluHisGlyGluGlu-ProGlyArgAspValLeuGluPr 1213
                    |||      |||:::      |||        |||            ||
Db    8308 GAAAAGGAGCCCGTGGATGAGCAAGAAGTGGAATCCGGAACCAAGG------------CC 8355

Qy    1213 oThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGl 1233
           |   ||||||   |||||||||:::                |||   |||      :::
Db    8356 ACTCCAGCAGAGTCTGACGGCCAGCCAA---------TTGATGAAATTGCCCCTGCA-AC 8405

Qy    1233 nGluGlyGluValAspTrpLeuAspGlyGluLysValLysGluGluGlnGluValPheVa 1253
               |||   :::|||                        :::|||:::          ||
Db    8406 ATCTGGACCCATCGATGAGGCTTCCACAGCCGCTCCCACAAAGGAGGAATCCACAACCGT 8465

Qy    1253 lHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGl 1273
           |   |||               |||:::       :::|||  |||:::
Db    8466 TGCTAGCGCA---------------GCTTCGCCTGCTGTTCACGATGATGAAATTAAGGA 8510

Qy    1273 yValAlaGlyCysGln------------GluLysGluSerThrGluValGlnSerLeuSe 1289
             |||      |||              ||||||||            |||
Db    8511 TGTCACCACTACGCAACCAGTTGCCGATGAGAAGGAAGTGGCTGCTCCTCAG-------- 8562

Qy    1289 rLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGl 1309
               :::|||   :::         ||||||      :::                |||
Db    8563 ----GACGAAACGAAGACATCCATTGATGTTTCCACCGAT------------TCACCCAC 8606

Qy    1309 uGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGl 1329
                     :::|||   :::   :::|||   ||||||               |||
Db    8607 TGCCCAAGATGATGAGAAGCAAGACAAAACCGAGGCTCCAGTGGCACCCACCACTGTTTC 8666

Qy    1329 yLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLe 1349
              |||      |||   |||         :::
Db    8667 ATCGCCG------ACTGCTGATTCTGCTGCTGATTCT----------------------- 8697

Qy    1349 uGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGl 1369
                          |||   |||          :::     :::|||:::       ::
Db    8698 -------------AGTACTCCCACAGTGGAGGTACCATCTCCTGTTGAAATCGACACGAA 8744

Qy    1369 nSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrVa 1389
           :   :::|||        ::::::||||||
Db    8745 GCCAATGGATGATATCATGTCCCAGACT-------------------------------- 8772

Qy    1389 lValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGl 1409
             :::            |||   :::               ||||||            ||
Db    8773 -ATTGCTCCTCATACCGCCGAT---------------GGTGCTGCGTCCACGTCCACTGA 8816

Qy    1409 uLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLe 1429
           |                           :::   |||   ||||||   |||   |||::
Db    8817 AGATGAAGACCAGGCTCCCGTTACTGTGTCTCCTCAGGATGCTGAAAAGACT---CCAGT 8873

Qy    1429 uGlyProGluSerGln---------------------------------------AlaGl 1436
           :   |||      |||                                       |||::

Db    8874 TTCCCCGGCTCCCCAGGATTCTGATAAGACTCCGAGCTCAGAGGCTCCCCAGGATGCCGA 8933

Qy    1436 uSerIleProIleIleValThrProAlaProGlu------------SerThrLeuHisPr 1452
           :   ||||||         ||||||         :::          :::|||:::   ||
Db    8934 CGAAATCCCAGCTACAGCGACACCTCTTGACGACAACAAGATCCCAGCCACTGTGGCTCC 8993

Qy    1452 oAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysPr 1472
           |               :::   |||:::                  |||::::::   :::
Db    8994 TCAGACTGATGATGGGGTTCCAGCTACCGCAGCTCCTCTTGATGAAGACAAGATCCAAAC 9053

Qy    1472 oAspAlaGlyPro---AspAlaAspGlyLysGluSerThrAlaIle-----------Gl 1487
              |||   |||      |||      :::      |||||||||             ||
Db    9054 TACTGCGGCTCCTCTTGATGAGGAGAAGATTCCATCCACAGCAGCTCCTCTTGATGATGA 9113

Qy    1487 uLysVal---LeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIl 1506
            |||:::               |||   :::   ::::::|||         |||   |||
Db    9114 AAAGATTCCTGCTCCAGTTAGCCCCGTTGTTTTCGATGTGGAGCCAAGCAGCGAGAAG-- 9171

Qy    1506 eValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGlu---ThrAlaPr 1525
                                  ||||||   ::::::            |||   |||   ||
Db    9172 -------------------CCTGCTGTATCTGAGTATGATGGTGAAGAGAGCACCGAACC 9212

Qy    1525 oGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGl 1545
           |      |||         ::::::   |||      |||
Db    9213 ACCTGTCCATGATGTGGAGACCAGCACTGATGAGCCTACT-------------------- 9252

Qy    1545 uProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnPr 1565
                                          ||||||||:::|||   |||         ||
Db    9253 ------------------------TCTGATGCCAAACTGAAGCCACCTACTTCTGCACC 9287

Qy    1565 oArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLe 1585
           |            :::               ||||||      |||
Db    9288 TGCTACGCCATCGGAA---TCTCCAGCCACGGAGGCGGAGATTGTGCCAGAAACTGCAGC 9344

Qy    1585 uProArgLeuGlnLeuLysAlaPro 1593
              |||   |||:::   :::   |||
Db    9345 CCCAGAGCTTGAAAAGGAGGTGCCA 9369
```

RESULT 11
ABL05592
ID   ABL05592 standard; cDNA; 18506 BP.
XX
AC   ABL05592;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster expressed polynucleotide SEQ ID NO 11258.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ss.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   P-PSDB; ABB61489.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX

```
PS   Claim 1; SEQ ID NO 11258; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 18506 BP; 5394 A; 4676 C; 4890 G; 3546 T; 0 other;

Alignment Scores:
Pred. No.:              1.36e-19      Length:        18506
Score:                  565.50        Matches:       412
Percent Similarity:     36.99%        Conservative:  282
Best Local Similarity:  21.96%        Mismatches:    737
Query Match:            7.00%         Indels:        448
DB:                     23            Gaps:          81

US-09-902-432-4 (1-1596) x ABL05592 (1-18506)

Qy       6 SerThrGluGlnArgSerProGluGlnProAlaGlySer-------------------- 18
              |||   |||          ||||||     |||   |||
Db   10770 TCACAAGAAGCTTCCCGGCCAGAATCTGAGGCAGAAAGCTTGAAAGATGCGGCAGCACCG 10829

Qy      19 ---------------------AspThrProSerGluLeuValLeuSerGlyHisGlyPro 31
                                :::::: :::|||   |||      |||      |||
Db   10830 TCGCAGGAGACCTCTCGTCCCGAATCTGTTACAGAAAGCGTTAAGGATGGAAAATCCCCA 10889

Qy      32 AlaAla------GluAlaSerGlyAlaAlaGly-----------------AspProAla 43
              |||      |||||||||         |||                   |||   |||
Db   10890 GTCGCATCCAAAGAAGCGTCCAGACCAGCATCCGTTGCAGAAAATGCCAAGGATTCGGCT 10949

Qy      44 AspAlaAspProAlaThrLys---------LeuProGln---LysAsnGlyGlnLeuSer 59
           |||               :::        |||||||||   |||   |||   :::
Db   10950 GATGAATCAAAGGAACAACGACCAGAGTCGTTGCCTCAAAGCAAGGCGGGAAGTATTAAA 11009

Qy      60 SerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGly 79
                                |||:::       :::       ::::::|||    ::::::
Db   11010 GAT---------------GAAAAATCACCTTTGGCCTCTAAAGATGAAGCTGAGAAGTCT 11054

Qy      80 GlnGluGluGlu------------------------------ValValAspGluAspVal 89
           :::||||||                                 :::|||   ::::::|||
Db   11055 AAGGAAGAATCAAGGCGAGAGTCGGTGGCCGAACAATTTCCTTTGGTCTCCAAGGAAGTT 11114

Qy      90 GlyGlnArgGluSer---------------GluAspValArgGluLysAspArgValGlu 104
              :::      |||                :::|||      ||||||   :::   |||
Db   11115 TCCAGACCAGCATCCGTTGCAGAAAGTGTTAAGGACGAAGCTGAGAAGTCTAAGGAAGAA 11174

Qy     105 ---------------GluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAsp 119
                          :::     |||:::         :::|||            ||||||
Db   11175 TCACCTTTAATGTCTAAGGAAGCTTCCAGACCAGCATCCGTTGCAGGAAGTGTTAAGGAT 11234

Qy     120 GlyGlnGluGluThrSerGlu------------------IleIleGluGlnIleProAla 133
                    |||::::::   |||                 :::      |||:::   |||
Db   11235 GAAGCTGAGAAGTCCAAGGAAGAATCTAGGCGAGAGTCAGTTGCTGAAAAATCACCTTTG 11294

Qy     134 SerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGly 153
                                                 |||||||||
Db   11295 CCCTCCAAAGAAGCTTCGAGACCAGCTTCCGTTGCAGAAAGT------------------ 11336

Qy     154 PheLysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGlu 173
                                                          ||||||:::      :::
Db   11337 ---------------------------------------------GTAAAGGATGAAGCTGAT 11354

Qy     174 LysSerAspThrValGlnLeuLeuThrValLysLysAspGluGly---------GluGly 190
           ||||||                                   |||:::|||         |||
Db   11355 AAGTCC--------------------------AAAGAAGAATCTAGGCGAGAGTCGGGG 11387

Qy     191 AlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGly 210
           ||||||  |||                        |||
Db   11388 GCCGAAAAATCT-----------------------CCTTTGGCCTCCAAGGAAGCTTCC 11423

Qy     211 GluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeu 230
                    ||||||              :::|||         ||||||   :::
Db   11424 AGACCAGCTTCCGTTGCAGAAAGTATTAAGGACGAAGCTGAGAAGTCTAAGGAAGAATCT 11483

Qy     231 LysGlnGluGlnSerSerThrGluIleProLeu---------GlnAlaGluSerAspGln 247
           ::::::|||     :::   :::   ||||||          :::|||
Db   11484 AGGCGAGAGTCGGTTGCTGAAAAATCTCCTTTGCCCTCCAAGGAAGCTTCGAGACCAACT 11543

Qy     248 AlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLys 267
           :::         |||      |||||||||      ||||||   ::::::|||         :::
Db   11544 TCCGTTGCAGAAAGTGTTAAGGATGAAGCT---GAGAAGTCTAAGGAAGAATCTAGAAGA 11600

Qy     268 SerProGluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhe 287
                 :::            ||||||:::   |||:::      |||
Db   11601 GACTCGGTGGCCGAAAAATCTCCTTTGGCGTCCAAGGAAGCTTCC--------------- 11645

Qy     288 PheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAsp 307
                                            :::      |||      :::|||   :::|||
Db   11646 ---------------------------AGACCAGCTTCCGTTGCAGAAAGTGTTCAGGAC 11678

Qy     308 AspLeuGluThrAla------GluLysArgLysGluGlnGluAlaGluLysValAspGlu 325
           :::   |||   :::         |||   |||:::|||      |||||||||
Db   11679 GAAGCTGAGAAGTCTAAGGAAGAATCTAGGCGAGAGTCGGTGGCCGAAAAATCTCCTTTG 11738

Qy     326 GluGluLysGluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAsp 345
                    ||||||   :::   |||||||||         :::         ::::::   :::
Db   11739 GCCTCCAAGGAAGCTTCCAGACCAGCTTCCGTTGCAGAAAGTATTAAGGACGAAGCTGAG 11798

Qy     346 GlnAlaArg---------------------------------------------LeuSer 350
           :::::::::                                                |||
Db   11799 AAGTCTAAGGAAGAATCTAGGCGAGAGTCGGGTGGCCGAAAAATCTCCTTTGGCCTCCAA 11858

Qy     351 AlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAsp---LeuGluAla 369
                    :::|||      :::|||||||||   :::   |||   |||      |||
Db   11859 GGAAGCTTCGAG---ACCCAACTTCCGTTGCA-GAAAGTGTTAAGGATGAAGCTGAGAAG 11914

Qy     370 SerSerGluGluLysCys--------------------AlaProLeuAlaThrGluVal 382
           |||   ||||||                          :::|||||||||:::
Db   11915 TCTAAGGAAGAATCTAGTCGAGACTCGGTGGCCGAAAAATCTCCCTTGGCGTCC------ 11968

Qy     383 PheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSerThrVal 402
                    |||      :::               |||||||||
Db   11969 ---------AAGGAAGCTTCCAGACCAGCTTCCGTTGCAGAAAGTGTTCAGGACGAAGCT 12019

Qy     403 GluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGlu 422
           ||||||::::::||||||                     ||||||         :::
Db   12020 GAGAAGTCTAAGGAAGAATCTAGGCGAGAGTCGGTGGCCGAAAAATCTCCTTTG------ 12073

Qy     423 GlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGlu 442
              :::   |||:::         |||         :::||||||      |||      |||
Db   12074 GCCTCCAAGGAAGCTTCCAGACCAGCTTCCGTTGCAGAA------AGTGTAAAGGATGAA 12127

Qy     443 AlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHis 462
           ||||||   ::::::|||      :::||||||         |||:::
Db   12128 GCTGAGAAGTCCAAAGAA------GAATCTAGGCGAGAGTCGGTGGCC------------ 12169

Qy     463 ThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHis---ProGluGly 481
              :::   :::   |||:::   :::|||   :::   |||                |||
Db   12170 ---GAAAAATCTCCTTTGGCCTCCAAGGAAGCTTCCAGACCAGCTTCCGTTGCAGAAAGT 12226

Qy     482 IleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGly------ 499
           :::      |||   |||               :::   |||   :::|||   |||
Db   12227 GTTAAGGATGAAGCTGAGAAGTCTAAGGAAGAATCTAGGCG-AGAGTCGGTGGCCCGAAA 12285

Qy     500 ---------SerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGly 516
                    |||   :::||||||                  |||:::|||:::
Db   12286 AATCTCCTTTGCCCTCCAAGGAAGCTTCCAGACCAACTTCCGTTGCAGAAAGTGTTAAGG 12345
```

```
Qy      517 LysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHis 536
Db    12346 ATGAAGCTGAGAAGTCTAAGGAAGAATCAAGGCGA-GAGTCGGTGGCCGAAAAATCTTCT 12404

Qy      537 IleHisThrGluSer---------ProGluSerAlaAspGluGlnLysGlyGluSerSer 553
Db    12405 TTGGCCTCCAAGAAAGCTTCCAGACCAGCTTCCGTTGCAGAAAGTGTTAAGGATGAAGCT 12464

Qy      554 AlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeu------ 571
Db    12465 GAGAAGTCTAAGGAAGAATCTAGGCGAGAGTCGGTGGCCGAAAAATCTCCTTTGGCCTCC 12524

Qy      572 ---GluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLys 590
Db    12525 AAGGAAGCTTCCAGACCAGCTTCCGTTGCAGAAAGTGTTAAGGACGAAGCTGAGAAGTCT 12584

Qy      591 ArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgVal 610
Db    12585 AAGGAAGAATCTAGGCGAGAGTCGGTGGCCGAAAAATCACCTTTGCCCTCCAAAGAAGCT 12644

Qy      611 ArgArgProSer-----------------------GluSerAspLysGluGluGlu--- 621
Db    12645 TCGAGACCAACTTCCGTTGCAGAAAGTGTAAAGGATGAAGCTGATAAGTCCAAAGAGGAA 12704

Qy      622 ------------------LeuGluLysValLysSerAlaThrLeu------SerSerThr 633
Db    12705 TCTAGGCGAGAGTCGGGGGCCGAAAAATCTCCTTTGGCCTCCATGGAAGCTTCCAGACCA 12764

Qy      634 AspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGlu--------- 650
Db    12765 ACTTCCGTTGCAGAAAGTGTTAAGGACGAAACTGAGAAGTCTAAGGAAGAATCTAGGCGA 12824

Qy      650 ------------------------------------------------------------ 650
Db    12825 GAGTCGGTGACCGAAAAATCTCCTTTGCCCTCCAAGGAAGCTTCCAGACCAACTTCCGTT 12884

Qy      651 -----------------------GlnLysProGluGluProLysArgArgValAspThr 662
Db    12885 GCAGAAAGTGTTAAGGATGAAGCTGAGAAGTCTAAGGAAGAATCTAGGCGA------GAG 12938

Qy      663 SerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAla 682
Db    12939 TCGGTGGCCGAAAAATCTCCTTTGGCCTCCAAGGAATCTTCCAGACCAGCTTCCGTTGCA 12998

Qy      683 SerSerSerAspAspGlu------------------------------------------ 688
Db    12999 GAAAGTATTAAGGACGAAGCTGAGGGGACTAAGCAGGAATCTAGGCGAGAGTCGATGCCC 13058

Qy      689 ------GlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLys 706
Db    13059 GAAAGCGGTAAGGCGGAGAGTATCAAGGGTGAT------CAATCTTCTTTAGCGTCCAAA 13112

Qy      707 AspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGln 726
Db    13113 GAA---ACGTCAAGACCCGATTCTGTTGTGGAAAGCGTTAAGGACGAAACTGAAAAGCCA 13169

Qy      727 GlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrp 746
Db    13170 GAAGGGTCAGCGATCGACAAGTCGCAGGTGGCATCCCGACCAGAGTCCGTGGCAGTA--- 13226

Qy      747 GluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLys 766
Db    13227 ---AGTGCTAAGGACGAGAAGTCACCGCTTCATTCGCGTCCTGAATCAGTTGCCGACAAG 13283

Qy      767 AlaGluAspSerSer--------------------------ValGluGlnLeuSerThr 777
Db    13284 TCGCCAGATGCCTCGAAGGAAGCATCTCGTTCTTTATCGGTTGCCGAAACCGCCTCATCA 13343

Qy      778 GluIleGluProSer---ArgGluGluSerTrpValSerIleLysLysPheIleProGly 796
Db    13344 CCCATTGAAGAAGGTCCACGATCCATTGCTGACCTCAGTCTGCCCCTTAACCTTACAGGT 13403

Qy      797 ArgArgLysLysArg-----------------------------AlaAspGlyLysGln 806
Db    13404 GAAGCTAAGGGCAAGCTTCCAACTCTTTCGAGTCCCATTGATGTGGCAGAAGGAGACTTT 13463

Qy      807 GluGlnAlaThrValGluAsp--------------------SerGlyProValGluIle 819
Db    13464 TTAGAAGTAAAGGCGGAATCCTCCCCGCGGCCAGCTGTACTTTCAAAACCAGCGGAGTTT 13523

Qy      820 AsnGluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaVal--- 838
Db    13524 TCGCAGCCCGATACTGGGCATACAGCCTCCACGCCCGTGGACGAGGCTTCTCCCGTTCTG 13583

Qy      839 ---GluArgGluLysMetGlu-----------------------AlaGlnGlyAsnThr 849
Db    13584 GAGGAAATCGAAGTAGTTGAACAGCATACCACTTCTGGAGTCGGGGCTACGGGCGCAACA 13643

Qy      850 GluLeuProGlnLeuLeuGly---------------AlaValTyrValSerGluGluLeu 864
Db    13644 GCAGAAACGGACCTACTTGACCTGACTGAAACCAAATCCGAGACTGTAACCAAGCAATCG 13703

Qy      865 SerLysThrLeuValHisThrValSer------------------ValAlaValIleAsp 878
Db    13704 GAGACAACGTTATTTGAAACCCTTACCTCGAAGGTTGAAAGTAAGGTGGAAGTACTGGAG 13763

Qy      879 GlyThrArgAlaValThrSerValGluGluArg---------SerProSerTrpIleSer 895
Db    13764 AGC------TCGGTGAAACAAGTTGAAGAGAAAGTCCAGACCTCAGTAAAGCAAGCAGAA 13817

Qy      896 AlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGluGlu 915
Db    13818 ACCACTGTGACCGATTCCTTGGAACAACTAACCAAGAAGAGCAGTGAGCAGTTGACAGAG 13877

Qy      916 Val---------------ThrGluLysAspIleIleAlaGluGluThrProValLeuThr 930
Db    13878 ATCAAATCAGTGCTAGACACCAACTTTGAAGAAGTAGCGAAAATCGTTGCCGATGTGGCC 13937

Qy      931 GlnThrLeuProGluGlyLysAspAlaHisAsp--------------AspMetValThr 945
Db    13938 AAGGTACTTAAGTCTGACAAGGACATTACAGACATCATACCCGACTTTGACGAACGGCAA 13997

Qy      946 SerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArg 965
Db    13998 CTGGAGGAGAAACTGAAGTCAACTGCCGATACAGAGGAGGAGTCCGACAAAAGTACCCGG 14057

Qy      966 ThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaVal 985
Db    14058 GACGAGAAATCCCTGGAAATCAGCGTCAAAGTAGAA--------------------ATC 14096

Qy      986 SerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGlu 1005
Db    14097 GAATCTGAGAAGTCATCGCCCGAT------CAGAAGTCCGGCCCCATTTCCATCGAGGAG 14150

Qy     1006 SerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeu--------- 1022
Db    14151 AAGGACAAAATCGAGCAATCAGAAAAAGCACAGTTGAGGCAGGGAATATTGACCAGCAGT 14210

Qy     1023 ------GlnAlaValAla---------AspLysValLysGluGluSerGlnValProAla 1037
Db    14211 CGGCCGGAGTCAGTTGCCAGCCAGCCGGAGTCGGTGCCCTCCCCTTCCCAGTCAGCAGCC 14270

Qy     1038 ThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db    14271 AGCCATGAACATAAGGAGGTGGAGCTATCCGAGAGCCATAAAGCTGAGAAGTCTTCTCGC 14330

Qy     1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db    14331 CCCGAGTCCGTTGCCAGCCAGGTCAGCGAAAAGGACATGAAGACATCGCGCCCAGCATCC 14390

Qy     1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db    14391 AGCACCAGCCAATTTAGCACCAAGGAAGGCGATGAG---------GAGACGACTGAGTCG 14441

Qy     1098 LeuGluValProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLys 1117
```

```
Db  14442 CTA-----------------------CTTCACTCGCTTACTACCACAGAGACAGTGGAG 14477

Qy   1118 LeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSer 1137
            :::|||:::   |||::::::            |||    ::::::|||   |||
Db  14478 ACGAAACAGATGGAAGAGAAATCGAGCTTTGAGTCTGTCTCCACTTCAGTCACTAAGTCC 14537

Qy   1138 -----------------------------------GluThrAsnGlySerThrProLeu 1145
                                             |||   :::   |||     |||
Db  14538 ACAGTCCTCTCCAGTCAGTCAACTGTTCAACTTCGAGAAGAGTCCACCAGCGAATCGTTG 14597

Qy   1146 AlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSer 1165
          :::   |||         :::   :::   :::   |||::::::   |||      :::
Db  14598 AGTAGCAGTCTCAAGGTGGAGGACAGCTCCCGACGCGAGTCGCTTTCCAGCTTGTTGGCG 14657

Qy   1166 LysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAla--- 1184
          :::                     :::   :::   |||:::   :::|||:::|||
Db  14658 GAGAAAGGAGGAATCGCTACCAACACTAGTTTGAAGGAGGACACTAGCGCCTCTGCATCT 14717

Qy   1185 --------------------GlnLysGluGluProSerThrLeuProAsnAsnValPro 1197
                              |||   ||||||   |||:::
Db  14718 CAGCTGGAGGAGCTGCTTGTGCAGTCCGAGGAATGCTCATCAGAGTCAATCGTTAGCGAG 14777

Qy   1198 AlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGlu 1217
             |||         ::::::      :::::::::   :::      :::   :::
Db  14778 ATTCAGACCAGCATCGCTCAGAAGTCAAATAAGGAAATCAAGGATGCTAGAGAAACCAAA 14837

Qy   1218 LeuThrAla---------AlaAlaValProValLeuAlaLysThrGluValGlyGlnGlu 1234
          :::|||:::                                |||   :::      :::|||
Db  14838 GTAACCAGCCAGTTTACCACTACCACCAGCAGCGCCACCAAAGACGATAGCTTAAAGGAA 14897

Qy   1235 GlyGluValAspTrpLeuAspGlyGluLysValLysGluGluGlnGluValPheValHis 1254
                      ::::::|||      ||||||:::         :::|||   |||
Db  14898 ACTGTGGCCGAGTTTCTGGCCACCGAGAAGATTGTTTCCGCCAAGGAAGCCTTC------ 14951

Qy   1255 SerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyVal 1274
                      :::   |||:::||||||
Db  14952 ---AGCACTGAAGCCACTAAGTCGGCCGAC------------------------------ 14978

Qy   1275 AlaGlyCysGlnGluLys---GluSerThrGluValGlnSerLeuSerLeuGluGluGly 1293
                |||   :::|||      ::::::   |||   |||   |||      :::
Db  14979 ---GACTGCTTGAAGAAGACCACCGCCAGCGCCGTCTCCAGCACATCCGCCTCCCAGCGA 15035

Qy   1294 GluMetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGlu 1313
             :::         ::::::   :::||||||:::            |||   ||||||
Db  15036 GCGCTATTCGTAGGCACGGACGAATCACGGAGAGAGTCTCTGCTCAGTCAGGCAAGTGAG 15095

Qy   1314 ------------GluGlyGluGlnGluThrAlaAlaProGluHisGlu-GlyThrTyrGl 1329
                            :::   |||         |||      :::   |||   ||
Db  15096 AGCCGCTTGACCCACAGTGATCCTGAAGATGAGGAACCCGCTGATGACGTCGACGAGCGG 15155

Qy   1329 yLysProValLeuThrLeuAspMetPro------------------------------ 1338
          |  |||||||||            :::|||
Db  15156 T---CCAGTGTTAAGGAGAGCCGTTCCAAGTCTATTGCTACCATCATGATGACCAGCATC 15212

Qy   1339 -------------------------------------SerSerGluArg-GlyLysAlaL 1346
                                               ||||||   |||      :::|||
Db  15213 TACAAACCTTCCGAGGATATGGAACCGATTTCAAAGCTCGTCGAGGAGGAGCACGAGCAC 15272

Qy   1346 euGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluV 1366
             |||      |||               |||      ||||||               :
Db  15273 GTGGAGGAGCTGGCTCAGGAGGTCACCTCTACCAGCAAGACCAC---AACTCTTTTGCAA 15329

Qy   1366 alGln-ValGlnSerLeuAspThrThrValThrGlnThr---------AlaGluAlaVal 1382
          ::|||   ||||||         ||||||   :::   :::         |||      |||
Db  15330 TCCAGCGAACAGAGCAGCACAACCACCAGTAGCACCAGCAAGACCGGTGCCTCAAGGGTG 15389

Qy   1383 GluLysValIleGluThrValValIleSerGluThrGlyGluSer--------------- 1397
          |||   :::         |||         :::|||   :::|||
Db  15390 GAGTCCATTACCCTCACACAGATGGATCAGCAAACTAGCCAGAGTCAGGGCGACCCTGCC 15449

Qy   1398 ------------ProGluCysValGlyAlaHisLeuLeuProAlaGluLysSer---Ser 1412
                      |||               |||   :::   |||      |||:::   |||
Db  15450 GATCGCAAGACACCGCCAACT------GCGCCAGTAAGTCCCGGCGTCAAAGCAATGAGC 15503

Qy   1413 AlaThrGlyGlyHisTrpThrLeuGlnHisAla---GluAspThrValProLeuGlyPro 1431
          :::||||||         ::::::         |||            |||      |||
Db  15504 TCGACAGGATCAGCCGGCTCTGTGATTGGAGCTGGAGCTGGAGCCGTAGCCGCTGGTGGA 15563

Qy   1432 GluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHis 1451
          :::   ::::::   :::         |||:::   :::         |||   |||
Db  15564 AAGTGTGAGTCGAGTGCCGCCAGCATCGTATCTTCATCGGGCCCGATGTCGCCC------ 15617

Qy   1452 ProAspLeuGlnGlyGluIleSer------------AlaSerGlnArgGlu--------- 1464
             |||:::   |||:::   |||                  |||:::   :::
Db  15618 AAGGACATCAGCGGCAAATCTAGCCCCGGTGCCCTCACCTCGGAGAGCCAGTCAATTCCC 15677

Qy   1465 ------------ArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
                      |||                  |||   |||::::::   |||
Db  15678 ACTCCGCTAGGACGGGAGTCGCATACGGACACCCCGGAATCCTCGCCA------------ 15725

Qy   1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
             |||   ::::::         :::|||   |||   |||   :::   ||||||::::::
Db  15726 AAGCCCACATCACCTTTCCCGCGCGTCAGCAAGGACGAGCTTAAGTCTCTGGAAATGCAG 15785

Qy   1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
                :::         ||||||::::::                        |||
Db  15786 GTATGTTTAAGCAGTTTTGTGCTGGATTTATTTGTCGTATTATATCCTATGTTTTTTAAC 15845

Qy   1521 ---ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValPro------ 1537
                   |||      :::   :::      |||         |||
Db  15846 CCATGTAGCACCACTCGCAGGAGCAGATGTTGGCAGGAGCAGCTGCTGCAGCCGGAGCTG 15905

Qy   1538 AlaCysArgLeu-AspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLy 1557
          :::   |||:::   ||||||                  ||||||::::::
Db  15906 AGTGTGAGGGTGACATTCCCGAGCTGCACGAACTGCGTGGACTAGAGTG----------- 15954

Qy   1557 sMetLysHisProValProGlnProArgGluAspLeuGln 1570
                |||      ||||||   ||||||      |||
Db  15955 -------CACCACAGCCCTCAG---CGGGAGCACCGACAA 15984

RESULT 12
ABL19989
ID   ABL19989 standard; DNA; 20448 BP.
XX
AC   ABL19989;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 11440.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 11440; 21pp + Sequence Listing; English.
XX
```

```
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 20448 BP; 7528 A; 4083 C; 4878 G; 3959 T; 0 other;

Alignment Scores:
Pred. No.:              3.9e-19         Length:         20448
Score:                  557.50          Matches:        396
Percent Similarity:     36.48%          Conservative:   300
Best Local Similarity:  20.75%          Mismatches:     725
Query Match:            6.91%           Indels:         487
DB:                     23              Gaps:           79

US-09-902-432-4 (1-1596) x ABL19989 (1-20448)

Qy      7 ThrGluGlnArgSerProGluGlnProAlaGly-------------------------- 17
Db   8965 ACTACAAAACGAATGTTGCGACGACCTGCTGGCGAAGGTGAAATAGAGATTATTGAAGTT 9024

Qy     18 ------SerAspThrPro---SerGluLeuValLeuSerGlyHisGlyProAlaAlaGlu 34
Db   9025 GTTAGGGATGATCAGCCTGAGGCAGAAATAACTATTGTGGAGTATGAGCCA--------- 9075

Qy     35 AlaSerGlyAlaAlaGlyAspProAlaAspAlaAspProAlaThrLysLeuProGlnLys 54
Db   9076 ------------------GAACCTGTGAACCAAGACGAAAAACCCAAAGAACCTAAGAAG 9117

Qy     55 AsnGlyGlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHis------Val 72
Db   9118 AAAACCAGG-----------------AAGGTTAAAAAGGATGATATACATGACTACATA 9159

Qy     73 Gln----------------------------------GluGluAsnGlnGluGlyGln 80
Db   9160 CAAAAGTTAATTGAGTTGGAAACACCAAAGACGGAGCTTGAAAAGTATGAGAAAATTGAA 9219

Qy     81 GluGluGluValValAspGlu---------------------AspValGlyGlnArgGlu 93
Db   9220 TTCGAACCAATAGTCAAGGACAAGCCATTGGATAGCCCGATTGACGTTTTGGACGAATCG 9279

Qy     94 SerGluAspValArgGluLysAspArg------------------------------- 102
Db   9280 CCAAAAGAAGTGCAGAAAAAAGATAAAAAGTCAAGGTCTACAAAAGTTCCGAATGAGGAA 9339

Qy    103 ------ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
Db   9340 ACTCCTGTTCAAGAACAATATGCAAAGGTCAATGTTGTCGAAGAAGAAGCTCCTGAGCAA 9399

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
Db   9400 CCCGAAATACCAGTGCAAATTCTAGAAGTTAAGCCCGTA---GAGGTAGATGTTAAAGAG 9456

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
Db   9457 GTTATAACTGAAGATGGCAAGCCGGTTCAAGAAAAGACTACTAAGCGAGTTTTAAAGAAG 9516

Qy    161 ValGly-----------------PheLysPheThrValLysLysAspLysAsnGluLys 174
Db   9517 ATTGGGCCCGAGGAACAAACCACTTTTAAGATTACAATG------ATTGAAAGCGAGGAC 9570

Qy    175 SerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSer 194
Db   9571 AACGATTCGGTT---------ACAGTTATTGTTGACGAAGAACCGGAAATAGCTTCTCCA 9621

Qy    195 ValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSer 214
Db   9622 CAGTCTATCGAAGAACATCCAGAACAGTCTAAGGAAAAGCTAGCTCCAAAGCCCAAGAAA 9681

Qy    215 ---------------LysGluSerGluLeu---------------------------- 219
Db   9682 ACTGTCAGAAAAGTTAAAAAGGACGACTTAAGCGATTACGTAAAGAAACTAATTGAAGAA 9741

Qy    220 ---------LysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSer 236
Db   9742 GAAATTCCTAAGGTTGACCTTGAGAAATATGAA---------AAGGTAGAAATGCCAGAA 9792

Qy    237 ThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGlu 256
Db   9793 AAACCAGTTAAACTAACTGTT---TCTGATTCCATTCCGGAAGAACCAAAACCTGACAAA 9849

Qy    257 GlyGlu-----------------GluLysGlnGluLysGluProThrLysSerProGlu 270
Db   9850 TCCCAACCAATATCAGTACTACCCGACACAACGAAACCCAAGAAAACGAAGACTCCCAAG 9909

Qy    271 SerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHis 290
Db   9910 ACTCCAAAGACAGAAGACACAGACCAGCAAGTTCCGGACGAACCGACAGAGACCACTGTA 9969

Qy    291 GlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGlu 310
Db   9970 GACACTACTGACATACCAGAGCTTACACCAACCCAAACTGCACAGCCCGAGGACACAGCC 10029

Qy    311 ThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGlu--------- 327
Db  10030 ACTGCACAAATTACACCAAGCGCACAAGAAGAGAAATCTACTCAAGATGACACTAAAGAT 10089

Qy    328 --------------------LysGluLysThrGluProAlaSerGluGluGlnGluPro 340
Db  10090 ACAATTCAGAAAACAGTTAAACATAAGAAAACAAAACCA------GACACACAAAAAAGC 10143

Qy    341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr-------------------- 353
Db  10144 GTTGAAACAAGTGAACTACCAGAGGTTCATAAAGACTATCAAATTAGCATCATCCACGAA 10203

Qy    354 ---------------GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
Db  10204 GAGCTTGTCGAGGAAGAGCAGCCCGAAAAGATTTTGGAGGTTAGAGTTATCGAT---GAA 10260

Qy    369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGlu--------- 385
Db  10261 GTCGCCGAGGTGGAAGAATCACAGCCCATTGTGGAAGAGGTAGAAGATGAGGAACCACAG 10320

Qy    385 ------------------------------------------------------------ 385
Db  10321 CCAGCTACAGAAGAAACTGTTGAGGACGTCACTAAGCCTAAGTCTAAAAAGAAGAAGGTG 10380

Qy    386 ---------LysMetGluAlaHisGlnGluValValAlaGluValHisValSerThrVal 402
Db  10381 GTTAAAAAGAAGACCGATGACCATGACGAACTCATCAAAAAGATGTTGGAGCAGGAAATA 10440

Qy    403 GluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGlu 422
Db  10441 GAGAAGACGGAACTGGAAAAATACGAGAAAATTGAATTTGATGTTCCCAAGAAACTCAAA 10500

Qy    423 GlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro------ 440
Db  10501 CCAGAATTCGCTGCTCTCGAACCAATAAAAATAGAGCGCAAAGAACAGAAGCCAACAAAG 10560

Qy    441 ---------------------------------------------------------Gln 441
Db  10561 GTGACTATTCTCGATGCCACTGATGTTCCTAAAACCGTGAAACTAAAACCGAGTAAACGT 10620

Qy    442 GluAlaGluProAlaGluGluLeu-----------------------MetLysSerArg 453
Db  10621 AAGGAGAAACCTGCCGAAGAACTGACCGTTCAACTTCCAAAGTTCAGACTTAAGGCCCGT 10680

Qy    454 GluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGluLys 473
```

```
Db  10681 ATGGTCTTGGTTGAA-----------------------------TATCCTCCAGCTCCC 10710

Qy    474 ThrLeuProLysHisProGlu---GlyIleValSerGluValGluMetLeuSer------ 490
Db  10711 CTTATTCCTAAAACCACTGATATTGGCGCCATTAAAGACAATGGCGAATTGTCTCGTAAC 10770

Qy    491 ---------SerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSer 507
Db  10771 ATAGAAGAGGCTGAGGAAATTCTGAAGTTCAAACCGCATAAGACCAAGAAAATT------ 10824

Qy    508 SerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGly 527
Db  10825 --------------------------AAAAAAATCAAG---------------------- 10836

Qy    528 AspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGlu 547
Db  10837 ------------------------------------GATGATTTAGAAAAGGTCGAGCTC 10860

Qy    548 GlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGlu 567
Db  10861 GAAAAGTACGAAAAATACATTAGTAGCGAGGAAGAACCAGAGGAAAAGACTCCTTATAAG 10920

Qy    568 LysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThr---------Thr 584
Db  10921 AAA---CCGGAAAAGGCACCAAAACCAGAGGAAAAACAAGAGGATGTGAAACTTAAGTTG 10977

Qy    585 SerAspGlyGluLysLys---ArgGluGlyIleThrProTrpAlaSerPheLysLysMet 603
Db  10978 GGCAAGGGTAAGAAAAAGCCAAAGGAGGAAGAAGCGCCC---------------GAAAAT 11022

Qy    604 ValThrProLysLysArgValArgArgProSerGluSerAspLysGlu------------ 619
Db  11023 GTTACTCTTAAAAACATTCCGCAAAAGCCACAAGAAGTTGAAGAAGAAGTTGAACTTAAA 11082

Qy    620 ---------GluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThr 636
Db  11083 CAAAAACCAAAGGAGGTTGAAATTGTTGAGGAACAAACAAAGAAACCCAAGGACGGAGAG 11142

Qy    637 Val--------------------SerGlu------------------MetGlnAspGlu 643
Db  11143 TTCGTGGTGGAGCCATTTGAACCATCTGAGTTTGATCGTCCTGAATATGTACCCGATGAA 11202

Qy    644 ValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgValAspThrSer 663
Db  11203 CTTGAGCAGATCGAACATCCAGAGATACCCGAGAAAGTTAAGAAGCCGTCTAAGACC--- 11259

Qy    664 ValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSer 683
Db  11260 ------------------------------------AAGTACAAGCCTAAAGATAAATCT 11283

Qy    684 SerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGlu 703
Db  11284 AAGTCAGAGCCGGAAACCATTGTATCGGAGATTGTTGCCGGAGTTCCAAAGGAGGAAGAG 11343

Qy    704 AlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAsp 723
Db  11344 GCAATTCCCGAACAGGATGTTAAGTTCCGTAAGCCCGAGAGAGACGCTCCGGAAGAGACT 11403

Qy    724 GlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyVal 743
Db  11404 GATTCTGAAATTAAGCTACGTCCTGTACCTCAGGCTTCTAAGGATGAAAAT--------- 11454

Qy    744 SerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeu 763
Db  11455 ---------CCTGACGAACAAGCGCTAGTGACGCCT------AAGGCAGAGGAGCCCATT 11499

Qy    764 GluGluLysAlaGluAspSerSerValGlu--------------------GlnLeuSer 776
Db  11500 CCGCAAGAAATCGAAGATAAGGCCATTGATGACGAAAAGAAACCGAAGAAATCTAAGCCC 11559

Qy    777 ThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIleProGly 796
Db  11560 AAAAAGGTGCAGCCGAAGGAGCAAGAG------------ATTGCAAAGGAAGAGCCC--- 11604

Qy    797 ArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyPro 816
Db  11605 ---GAAGAATTTGAGGTATCGGTGAAAGAGGAAGAAGCTCTAGTGGATAAA------CCC 11655

Qy    817 ValGluIleAsnGlu---AspAspProAsnVal------------ProAlaValValPro 831
Db  11656 ATTGAAATCGAAAAACCGAAAGATGTTAAGGTGAAAGAGAAGAAACCCAAGGAAGCTCCT 11715

Qy    832 LeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGluLeu 851
Db  11716 GTGTCCGAAGTAGTAGTTATTGAAGAAGAGCCGAAACCAGAAGAAGTACCCGAGGAAATA 11775

Qy    852 Pro---------GlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeu 868
Db  11776 CCTGTGGAGTACAAGATAACAACTACAGTCTTGGAACCAGAGGATGCACCAAAGGAA--- 11832

Qy    869 ValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer------Val 886
Db  11833 ---CATCAAGTCAAAGTCATCGATTTTGACGAGCGTCAAGAAACTACGGAGGAGGTAATT 11889

Qy    887 GluGluArg--------------------------------SerProSerTrpIleSer 895
Db  11890 GAAGAAAAGGTCGTGACGCGCAAGAAGAAACCGAAACCACAGCAACCTGAAGAGTTCGAG 11949

Qy    896 AlaSerValThrGluProLeuGluHis-----------------ThrAlaGlyGluAla 909
Db  11950 GTCACCTTAAAGGAGCCAAAAGAAGAACAAATACAACCGGATGTAGTTTCAGCTGAAATC 12009

Qy    910 MetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProValLeu 929
Db  12010 TCGTTGCCAATTGAGGAGCCCGAACAGAAGCCAGAACAGTACGAGGTCGAACTTAAAATA 12069

Qy    930 ThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAsp 949
Db  12070 ACCCAAACTACACCAGAA------GAGCCCAATGATGTGCAAATCGCAGTCAAGGAAAAG 12123

Qy    950 PheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluVal 969
Db  12124 GTCAAGACCAAACCTGTTAAGAAAGTCAAGGAGGATAAAATTGTCGTGGTTGAGGCAGAA 12183

Qy    970 ThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGln------ 987
Db  12184 GAAGAGAAGCAACCAGTTGAGGAAACAATCGTGGAGGTTGAAAAGCAGGAAGAAAAGAAG 12243

Qy    988 LeuThrAspSerProAspThrThrGlu---GluAlaThrProValGlnGluValGluSer 1006
Db  12244 AAGTCGGAGAAACCAAAATCTTACGAGTTCAAAATTTCAGAGACGCAATCTATTGAAGAA 12303

Qy   1007 GlyValLeuAspThrGluGluGlu------GluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db  12304 AAGCCAATAGAAGTGGCTGAAGAAGCTCCTGAGGAAACACCAAAGGTTGTAGAGAAAAAG 12363

Qy   1025 ValAlaAspLys--------------------ValLysGlu------------------ 1031
Db  12364 GTTGCCGAGAAATTCGATTCCTACGAATTCACTCTAAAAGAAACTGATGAAGAAAAGGTA 12423

Qy   1032 --------------------GluSerGlnValProAlaThrGlnThrValGlnArgThr 1044
Db  12424 ATCACAGTTGACGACCAGCCCGAGGAAGAAGCCCCT---GTAGAGGTTGTATTCAAGAAA 12480

Qy   1045 GlySerLysAlaLeuGluLysValGluGlu--------------------------Val 1055
Db  12481 AAACCCAAAGAACCTGAAGCAGTTGAGGCCGAGTTTGTAATGACTGAACCCAAGATTGTG 12540

Qy   1056 GluGluAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyPro 1075
Db  12541 GAAGAAACGAGCGTGGAAACTGCAATCAAACAGAAAAAGACGAAAAAACCGAAGAAGGAC 12600

Qy   1076 ValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrPro 1095
Db  12601 GAAGAGGAAGCTCAATTAGCAATCAAGGTTGTCGAAAGTGAAGCCCCGGTTGCCGAAGAA 12660
```

```
Qy     1096 GluSerLeuGluValProGluValThrAlaAspValAspHisValAlaThrCysGlnVal 1115
Db    12661 GTGTTCAGTGAAGCTCCTGAGAGTAAGATTGTAGAGGAGGAAGTTATTGCAGAAGAAAAG 12720

Qy     1116 IleLys------LeuGlnGlnLeuMetGluGlnAlaValAlaProGlu------SerSer 1131
Db    12721 CCTAAAGAGTTCACAATTCGCGTTTCTGAATCAGAACCCAAGCCCGAGGAGCCAAGTGTC 12780

Qy     1132 GluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp------------ 1147
Db    12781 GAGCAGTTCACAGTTAAAAAACGCAAGCCTTCTGTTACTTTCGCCGACGAACCTGCCACT 12840

Qy     1148 ------------------SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAsp 1161
Db    12841 GAAATCGTTATCAAGGAAAGCAAGCCCGCTGAAGTTGTCACTGAAGACGCTCATATCAAG 12900

Qy     1162 SerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAla 1181
Db    12901 ACCAAGAAGCCAAAG--------------------AAGAAGGTAACGGATGTGGAAGCT 12939

Qy     1182 AlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGlu 1201
Db    12940 GAGGAGCTAAAGATTAAGATCACTGAGGAAGTTCCACAGGAAATTCCCATATTGGAGGAG 12999

Qy     1202 HisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAla 1221
Db    13000 GTCTCGGAGGAA---------GAGGTGATCACTGAAACCAAGAAG-----------ACA 13038

Qy     1222 AlaValProValLeuAlaLysThr---GluValGlyGlnGluGlyGluValAspTrpLeu 1240
Db    13039 GCGCCTGTTGTTGAAGAGAAGACATACAAAATTGGTATCAAGGAAACGGAGCCGGAAAAG 13098

Qy     1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
Db    13099 CCTGCCGAAGCAATTGTCGAGGAGGAGGAGCCGGTGGTAACAGAGCCAATTGAGGAAGCA 13158

Qy     1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
Db    13159 CCGAAACCGGAAGTCTTTGAGGAACACAAGGTTAGAGTCATT------------------ 13200

Qy     1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
Db    13201 ---------GAGGAAACACCTCGGGAATTGGTCGAGGAAGTCATTGAGGAGGAGGTGAAG 13251

Qy     1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
Db    13252 GTAATTCGCAGAAAGAAGCCTAAGCCAGAAATTAAAGAGGAACCCGAGGCGGAAGTTACT 13311

Qy     1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
Db    13312 GTT--------------TCTACTCCAAAGCCTGTC------------------------ 13332

Qy     1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
Db    13333 GAAGAAGTGGAAGCTACGTCCAGTATTGCTGTTATTCCAGAACAACCGACTGAGGAGGAA 13392

Qy     1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
Db    13393 GCCGCTGATCTCAAGATAACAATCATTGAAGAAGAA---------ACACCACCACAAGAA 13443

Qy     1381 AlaValGluLysValIleGluThrValValIleSerGlu--------------------- 1393
Db    13444 CTAGTTCAGGAAATTGAAGAAATTGAAATCGTTGAGGAACCAAAGGCCCCTGAAGAACAA 13503

Qy     1394 ---------------------ThrGlyGluSerProGluCysValGlyAlaHisLeuLeu 1406
Db    13504 CCAACGGATTTCACCTTCGCCACAAAGGATTCCGAAAAGAAGCCGACTGTGGAAGAGCTG 13563

Qy     1407 ProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThr 1426
Db    13564 CCCGAGGAACAGGTC---------------------ACGATTCAAAAAAAGAAAAAGAAG 13602

Qy     1427 ValProLeuGly---------ProGluSerGlnAlaGluSerIleProIleIleValThr 1443
Db    13603 GCACCAGTACCAGAGGTAGTTGAAGAACCCGAAGCTGAGTTTTGGTTAAGCCAAAGACG 13662

Qy     1444 ProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArg 1463
Db    13663 CCTGTCCAAGAAGTTACT------------GAAGAAGCTAAGATAACGAAGTCTAAGAAA 13710

Qy     1464 GluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSer 1483
Db    13711 CCTGTAAAGGAAGAAGAG-----------------------GCCGCGGCAGAGCTTAAA 13746

Qy     1484 ThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeu------------ 1499
Db    13747 GTAACGATCACTGAAGAAATTCCCACTGAACCAGAAGTTCAAGAAATTATCGAGGAGATC 13806

Qy     1500 ------------GluSerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaVal 1515
Db    13807 GAAGAGATCGAGGAAGAAAAACCCGCAGAGTATGTGATCGAAGTTAAGGAAAGCCAACCA 13866

Qy     1516 AspGlnPheAlaArgThrGluThr------------------------AlaProGluThr 1527
Db    13867 GAAGCTGTAGAAGATAAGGAAGTTAGCTTACCGAAAAAGAAACCTAAGGCTCCAATAGTT 13926

Qy     1528 HisAlaTyrAspSerGln---ThrGlnValProAlaCysArgLeuAspSerArgGluPro 1546
Db    13927 GAAGAGCCAGAAGCTGAAATAACTTTGAAGCCAAAGGTAAAGTCGGAAGAAGTTCAGGAA 13986

Qy     1547 AsnArgCysTrpThrLysMetLysAspAlaLysMet------------------------ 1558
Db    13987 GAAGCAAAGATAGTGAAGAAAAAGCCAAAGAAAATCGACGAAGTTGCAGTAGCCGACGAG 14046

Qy     1559 ------------------LysHisProValProGlnProArgGluAspLeuGlnValLeu 1572
Db    14047 CTTACGGTCAAGGTTGAAGAAGAGGTTGTCCCAGAGCCGATTGTTGAAGAAGAAGTGATA 14106

Qy     1573 ThrValLeuGluAlaTrpAlaGlnProArgLysCysLeuProArg--------------- 1587
Db    14107 GAAGAGTTCGAAATTAAAAAGAAACCAAAAGAACCGGAGCCAGAGGACATTGTCGATGCT 14166

Qy     1588 ------LeuGlnLeuLysAlaPro 1593
Db    14167 GCGATTGTGAAACTAAAGAAACCA 14190
```

RESULT 13
ABL19988/c

```
ID   ABL19988 standard; DNA; 24971 BP.
XX
AC   ABL19988;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 11437.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
```

```
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 11437; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 24971 BP; 5336 A; 5636 C; 4884 G; 9115 T; 0 other;

Alignment Scores:
Pred. No.:              5.46e-18      Length:         24971
Score:                  537.00        Matches:        350
Percent Similarity:     35.39%        Conservative:   252
Best Local Similarity:  20.58%        Mismatches:     625
Query Match:            6.65%         Indels:         475
DB:                     23            Gaps:           66

US-09-902-432-4 (1-1596) x ABL19988 (1-24971)

Qy        2 GlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThrPro 21
Db    11785 GGTAAGAAAAAGCCAAAGGAGGAAGAAGCGCCCGAAAATGTTACTCTTAAAAACATTCCG 11726

Qy       22 SerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAsp 41
Db    11725 CAAAAG------------------------------------------------------ 11720

Qy       42 ProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerVal 61
Db    11719 CCACAAGAAGTTGAAGAAGAAGTTGAACTTAAACAAAAACCAAAGGAGGTTGAAATTGTT 11660

Qy       62 AsnGlyValAlaGluGln---------GlyAspValHisValGlnGlu---GluAsnGln 77
Db    11659 GAGGAACAAACAAAGAAACCCAAGGACGGAGAGTTCGTGGTGGAGCCATTTGAACCATCT 11600

Qy       78 GluGlyGlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspVal 97
Db    11599 GAGTTTGATCGTCCTGAATATGTACCCGATGAACTTGAGCAGATCGAACATCCAGAGATA 11540

Qy       98 ArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThr 117
Db    11539 CCCGAGAAA------GTTAAGAAGCCGTCTAAGACCAAGTACAAGCCTAAAGATAAATCT 11486

Qy      118 LysAspGlyGlnGluGlu---ThrSerGluIleIleGluGlnIleProAlaSerGluAsn 136
Db    11485 AAGTCAGAGCCGGAAACCATTGTATCGGAGATTGTTGCCGGAGTTCCAAAGGAGGAAGAG 11426

Qy      137 AsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLys 156
Db    11425 GCAATTCCCGAA---------------------------CAGGATGTTAAGTTCCGTAAG 11393

Qy      157 ValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAsp 176
Db    11392 ------------------------------CCCGAGAGAGACGCTCCGGAAGAGACTGAT 11363

Qy      177 Thr---ValGlnLeuLeuThrVal---------LysLysAspGluGlyGluGlyAlaGlu 192
Db    11362 TCTGAAATTAAGCTACGTCCTGTACCTCAGGCTTCTAAGGATGAAAATCCTGACGAACAA 11303

Qy      193 AlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSer 212
Db    11302 GCGCTAGTG--------------------------------------------------- 11294

Qy      213 AlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGln 232
Db    11293 ------------------------------ACGCCTAAGGCAGAGGAGCCCATTCCGCAA 11264

Qy      233 GluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGlu 252
Db    11263 GAA---------------------------------------------ATCGAAGATAAG 11249

Qy      253 AlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerPro 272
Db    11248 GCCATTGATGAC---------------GAAAAGAAACCGAAGAAATCTAAGCCCAAAAAG 11204

Qy      273 SerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrp 292
Db    11203 GTGCAGCCG--------------------------------------------------- 11195

Qy      293 AlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGluThrAla 312
Db    11194 ------------AAGGAGCAAGAGATTGCAAAGGAAGAGCCCGAAGAATTTGAGGTATCG 11147

Qy      313 GluLysArgLysGluGlnGluAlaGluLysValAspGlu--------------------- 325
Db    11146 GTGAAAGAGGAAGAAGCTCTAGTGGATAAACCCATTGAAATCGAAAAACCGAAAGATGTT 11087

Qy      326 ---------GluGluLysGluLysThrGluProAlaSerGlu------------GluGln 338
Db    11086 AAGGTGAAAGAGAAGAAACCCAAGGAAGCTCCTGTGTCCGAAGTAGTAGTTATTGAAGAA 11027

Qy      339 GluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr--------------- 353
Db    11026 GAGCCGAAACCAGAAGAAGTACCCGAGGAAATACCTGTGGAGTACAAGATAACAACTACA 10967

Qy      354 ------GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSer 371
Db    10966 GTCTTGGAACCAGAGGATGCACCAAAGGAACATCAAGTCAAAGTCATCGAT--------- 10916

Qy      372 GluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln 391
Db    10915 ---------------------------------TTTGACGAGCGTCAAGAAACTACGGAG 10889

Qy      392 GluValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGly 411
Db    10888 GAGGTAATTGAAGAAAAGGTCGTGACGCGCAAGAAGAAACCGAAACCACAGCAACCTGAA 10829

Qy      412 GlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGlu 431
Db    10828 GAGTTCGAGGTC---------ACCTTAAAGGAGCCAAAAGAAGAACAAATACAACCGGAT 10778

Qy      432 ------------------------LysLeuAlaGluProGlnGluValProGlnGluAla 443
Db    10777 GTAGTTTCAGCTGAAATCTCGTTGCCAATTGAGGAGCCCGAACAGAAGCCAGAACAGTAC 10718

Qy      444 GluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThr 463
Db    10717 GAGGTC------------------------------------------------GAACTT 10706

Qy      464 GlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleVal 483
Db    10705 AAAATAACCCAAACTACACCAGAAGAG---------CCCAATGATGTGCAAATCGCAGTC 10655

Qy      484 SerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLys 503
Db    10654 AAGGAA------------------------------AAGGTCAAGACCAAACCTGTTAAG 10625

Qy      504 LysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArg 523
Db    10624 AAAGTC---AAGGAGGATAAAATTGTCGTGGTTGAGGCAGAAGAAGAGAAG--------- 10577

Qy      524 GlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGlu 543
Db    10576 ------------------CAACCAGTTGAGGAAACAATCGTGGAGGTTGAAAAGCAGGAA 10535
```

```
Qy    544 SerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThr 563
Db  10534 ------GAAAAGAAGAAGTCGGAGAAACCAAAATCTTACGAGTTCAAAATTTCAGAGACG 10481

Qy    564 ThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGlyThr 583
Db  10480 CAATCTATTGAAGAAAAGCCAATAGAAGTGGCTGAAGAAGCTCCTGAGGAA--------- 10430

Qy    584 ThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMet 603
Db  10430 ------------------------------------------------------------ 10430

Qy    604 ValThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeuGlu 623
Db  10429 ---ACACCAAAGGTTGTAGAGAAAAAGGTTGCCGAGAAATTCGATTCCTACGAATTC--- 10376

Qy    624 LysValLysSerAlaThrLeuSerSerThrAspSer------------ThrValSerGlu 639
Db  10375 ---------------ACTCTAAAAGAAACTGATGAAGAAAAGGTAATCACAGTTGACGAC 10331

Qy    640 MetGlnAspGluValLysThrValGlyGlu------GluGlnLysProGluGluProLys 657
Db  10330 CAGCCCGAGGAAGAAGCCCCTGTAGAGGTTGTATTCAAGAAAAAACCCAAAGAACCTGAA 10271

Qy    658 ArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLys 677
Db  10270 GCAGTTGAGGCCGAGTTTGTAATGACTGAACCCAAGATTGTGGAAGAAACGAGCGTGGAA 10211

Qy    678 ArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAsp 697
Db  10210 ACTGCAATCAAA------------------------------------------------ 10199

Qy    698 SerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAla 717
Db  10198 CAGAAAAAGACGAAAAAACCGAAGAAGGACGAAGAGGAAGCTCAATTAGCAATCAAGGTT 10139

Qy    718 SerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerPro 737
Db  10138 GTCGAAAGTGAAGCCCCGGTTGCCGAAGAAGTGTTCAGTGAAGCTCCTGAGAGTAAGATT 10079

Qy    738 SerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLys 757
Db  10078 GTAGAGGAGGAAGTTATTGCA--------------------------------------- 10058

Qy    758 LysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThr 777
Db  10057 ------------------GAAGAAAAGCCTAAAGAGTTCACAATTCGC------GTTTCT 10022

Qy    778 GluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArg 797
Db  10021 GAATCAGAACCCAAGCCCGAGGAGCCA------AGTGTCGAGCAGTTCACAGTTAAAAAA 9968

Qy    798 ArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThr---------ValGluAspSer 814
Db   9967 CGCAAGCCTTCTGTTACTTTCGCCGACGAACCTGCCACTGAAATCGTTATCAAGGAAAGC 9908

Qy    815 GlyProValGluIleAsnGluAspAspProAsnValProAlaValValProLeuSerGlu 834
Db   9907 AAGCCCGCTGAAGTTGTCACTGAAGACGCTCATATCAAGACCAAGAAGCCAAAGAAGAAG 9848

Qy    835 TyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeu 854
Db   9847 GTAACGGATGTGGAAGCTGAGGAGCTAAAGATTAAGATCACTGAGGAAGTTCCACAG--- 9791

Qy    855 LeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrValSerVal 874
Db   9790 ------GAAATTCCCATATTGGAGGAGGTCTCGGAGGAA--------------------- 9758

Qy    875 AlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIle 894
Db   9757 GAGGTGATCACTGAAACCAAGAAGACAGCGCCTGTTGTTGAAGAGAAGACATACAAAATT 9698

Qy    895 SerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGlu 914
Db   9697 GGTATCAAGGAAACGGAGCCGGAAAAG-----------------------CCTGCCGAA 9662

Qy    915 GluValThrGluLysAspIleIleAlaGluGluThrProValLeuThrGlnThrLeu--- 933
Db   9661 GCAATTGTCGAG---------------GAGGAGGAGCCGGTGGTAACAGAGCCAATTGAG 9617

Qy    934 ------------ProGluGlyLysAspAlaHis--------------------------- 940
Db   9616 GAAGCACCGAAACCGGAAGTCTTTGAGGAACACAAGGTTAGAGTCATTGAGGAAACACCT 9557

Qy    941 ------------AspAspMetValThrSerGluValAspPheThrSerGluAlaValThr 956
Db   9556 CGGGAATTGGTCGAGGAAGTCATTGAGGAGGAGGTGAAGGTAATTCGCAGAAAGAAGCCT 9497

Qy    957 AlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGlu 976
Db   9496 AAGCCAGAAATTAAAGAGGAACCCGAGGCGGAAGTTACTGTTTCTACTCCAAAGCCTGTC 9437

Qy    977 GluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGlu 996
Db   9436 GAAGAAGTGGAAGCTACGTCCAGTATTGCTGTTATTCCAGAACAACCG---ACTGAGGAG 9380

Qy    997 GluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluArg 1016
Db   9379 GAAGCCGCTGATCTCAAG------------------ATAACAATCATTGAAGAAGAAACA 9338

Qy   1017 GlnThrGlnAlaIleLeuGlnAlaValAlaAsp---LysValLysGluGluSerGlnVal 1035
Db   9337 CCACCACAAGAACTAGTTCAGGAAATTGAAGAAATTGAAATCGTTGAGGAACCAAAGGCC 9278

Qy   1036 ProAlaThrGlnThrValGlnArgThr---GlySerLysAlaLeuGluLys--------- 1051
Db   9277 CCTGAAGAACAACCAACGGATTTCACCTTCGCCACAAAGGATTCCGAAAAGAAGCCGACT 9218

Qy   1052 ValGluGluValGluGluAspSerGluValLeuAlaSerGluLysGluLysAsp------ 1069
Db   9217 GTGGAAGAGCTGCCCGAGGAACAGGTCACGATTCAAAAAAAGAAAAAGAAGGCACCAGTA 9158

Qy   1070 ------------------------------------ValMetProLysGlyProValGln 1077
Db   9157 CCAGAGGTAGTTGAAGAACCCGAAGCTGAGTTTTTGGTTAAGCCAAAGACGCCTGTCCAA 9098

Qy   1078 ----------------------------------------------------GluAla 1079
Db   9097 GAAGTTACTGAAGAAGCTAAGATAACGAAGTCTAAGAAACCTGTAAAGGAAGAAGAGGCC 9038

Qy   1080 GlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGlu 1099
Db   9037 GCGGCAGAGCTTAAAGTAACGATCACTGAAGAAATTCCCACTGAACCAGAAGTTCAAGAA 8978

Qy   1100 ValProGluValThrAlaAspValAspHisValAlaThrCysGln---ValIleLysLeu 1118
Db   8977 ATTATCGAGGAGATCGAAGAGATCGAGGAAGAAAAACCCGCAGAGTATGTGATCGAAGTT 8918

Qy   1119 GlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGlu 1138
Db   8917 AAGGAAAGCCAACCAGAAGCTGTA---------GAAGATAAGGAAGTTAGCTTACCGAAA 8867

Qy   1139 ThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGlu 1158
Db   8866 AAGAAACCTAAGGCTCCAATAGTTGAAGAGCCAGAAGCTGAAATAACTTTGAAGCCAAAG 8807

Qy   1159 ThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGlu 1178
Db   8806 GTAAAGTCGGAAGAAGTTCAGGAAGAAGCAAAGATAGTGAAGAAAAAGCCAAAGAAAATC 8747

Qy   1179 GluGluAlaAlaThrAlaGln------------LysGluGluProSerThrLeuProAsn 1194
Db   8746 GACGAAGTTGCAGTAGCCGACGAGCTTACGGTCAAGGTTGAAGAAGAGGTTGTCCCAGAG 8687

Qy   1195 AsnValProAlaGlnGluGluHisGlyGlu---GluProGlyArgAspValLeuGluPro 1213
```

```
Db   8686 CCGATTGTTGAAGAAGAAGTGATAGAAGAGTTCGAAATTAAAAAGAAACCAAAAGAACCG 8627

Qy   1214 ThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu-------------------- 1226
               ::::::::    ||||||:::     |||
Db   8626 GAGCCAGAGGACATTGTCGATGCTGCGATTGTGAAACTAAAGAAACCAGAGCCAGTAGAT 8567

Qy   1227 ---------------------------------------AlaLysThrGluValGlyGln 1233
                                                  |||||||||||||   :::
Db   8566 GCTGATGAAGTTGTGGCTGAAGTTACCCTTAAGCCAAAGGCTAAAACGGAAGTTACAGAG 8507

Qy   1234 GluGly---------GluValAspTrpLeuAspGlyGluLysValLysGluGlu------ 1248
           |||            :::|||          :::||||||  ||||||
Db   8506 GAAGAATTTTCTGTTGATGTTAAGCTGCCAAAGGAAAAGAAGGTTACAGAAGAAACTTCA 8447

Qy   1249 ---GlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyr 1267
             |||  |||   :::                  ||||||       :::
Db   8446 GACCAAACTGTACAGCTCAAGAAAAAGAAGAAGCCGCAGAAGCCGGTGGAAGAGGCTGCG 8387

Qy   1268 AspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSer 1287
           |||          :::                ::::::     |||    :::
Db   8386 GACGAGCTTAAGCTA------------------CAGCAAACTGTTGTCGAGGAAAGACCT 8345

Qy   1288 LeuSerLeuGluGluGlyGlu---MetGluThrAspValGluLysGluLysArgGluThr 1306
          :::   :::||||||  |||    :::|||       |||       |||:::
Db   8344 GTGGAAATTGAAGAAGAAGAGATCATTGAAGAAGCAGTCGTCATTCGCAAAAAGCCAAAG 8285

Qy   1307 LysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGly 1326
           ||||||   :::   :::   |||      :::||||||      |||
Db   8284 AAGCCCTTCGAGCCGACGGTTGAAGATCTCGAAGAAACA---------GAGTTCAGTCTC 8234

Qy   1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeu 1346
          ::::::    ||||||    |||::::::             |||          |||
Db   8233 TCCTTCAAAAAGCCC---CATACCATTAATGAGGGTGTGGAGGAGGCCGCCACAGTTCTG 8177

Qy   1347 GlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluVal 1366
                            |||:::     |||:::    |||
Db   8176 AAAAAACGACCAGTCAAGCCCACTACTCTGGATGAAGCTGCAGCT--------------- 8132

Qy   1367 GlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGlu------Lys 1384
                         :::   ::::::   :::       |||     |||        :::
Db   8131 ---------------GAACTTTCTATTAAGCGTCAGGAAGAAGAATACGAAGAGGGCGAG 8087

Qy   1385 ValIleGluThrValValIleSerGlu--------------------------ThrGly 1395
             ||||||     |||:::|||:::                            |||
Db   8086 GATATCGAAGAGTTCGTCGTTAGTCAGCAAAGAAAGCCGAAGCCTTTGCAAATCACAGAA 8027

Qy   1396 GluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGly 1415
           |||      |||                   |||         |||
Db   8026 GAGGATGAGGAGGCCTACACAGTGAAGAAACTAAAGCGTCGAAAGCAA------------ 7979

Qy   1416 GlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAla 1435
                                              |||   :::            :::
Db   7978 -----------------------------------GTAGACATTCCCGAATATGCTGATGTT 7952

Qy   1436 GluSerIleProIleIleValThrProAlaProGluSerThrLeuHisProAspLeuGln 1455
           |||::::::                  |||        :::            |||:::
Db   7951 GAGAACGTCACATTC---------CGTGCGAGAAGTACCAAGACCAAGGAAGATGTCGAT 7901

Qy   1456 GlyGluIleSerAlaSerGlnArgGluArgSerGluGluGlu------------------ 1469
             |||   :::   :::               :::|||||||||
Db   7900 CAAGAATTCAACATTGCCCTGGACTCGTATGCCGAGGAGGAAATTTCCATGTCTGGAAAG 7841

Qy   1470 ---------AspLysPro------------------AspAlaGlyProAspAla------ 1478
                     ||||||                    :::|||      :::|||
Db   7840 GTGAAGTTAAAGAAGCCTATCAAGAAGACCTTCTCTGAAGCAGCCGATGAAGCCAAGATC 7781

Qy   1479 ------------------AspGlyLysGluSerThrAlaIleGluLysValLeuLysAla 1492
                            |||      |||      ||||||::::::
Db   7780 AAGATCATTCAGGACTTTGACGATGGTGAAGAACCCATTATCGAAGAAATTCGAGACGAT 7721

Qy   1493 GluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnValIleGln 1512
           |||.     |||   |||:::|||                  |||         :::
Db   7720 GAGGATACCATCGATGAAGTCGAGGAACCAGAAGAGTACTTTGTTGAGGAATTGCCACCA 7661

Qy   1513 ThrAlaValAspGlnPheAlaArgThrGluThrAlaProGluThrHisAlaTyrAspSer 1532
               ||||||       |||       :::   ||||||:::|||       :::   |||
Db   7660 GACGAAGTGGACTT-CAAGCTTAAGCCAAAGAAGCACCCCAAACCAGCATATTCTGTTCA 7602

Qy   1533 Gln 1533
               :::
Db   7601 AGA 7599

RESULT 14
ABL28640
ID   ABL28640 standard; DNA; 24789 BP.
XX
AC   ABL28640;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 37393.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC,  Adams M,  Li PWD,  Myers EW;
XX
DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Claim 1; SEQ ID NO 37393; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 24789 BP; 7654 A; 5448 C; 6281 G; 5406 T; 0 other;

Alignment Scores:
Pred. No.:              1.17e-17      Length:        24789
Score:                  530.50        Matches:       389
Percent Similarity:     36.85%        Conservative:  309
Best Local Similarity:  20.54%        Mismatches:    639
Query Match:            6.57%         Indels:        559
DB:                     23            Gaps:          79

US-09-902-432-4 (1-1596) x ABL28640 (1-24789)

Qy      8 GluGlnArgSerProGluGlnProAlaGlySerAspThrProSerGluLeuValLeuSer 27
            ||||||:::|||   ::::::|||       |||:::       :::|||   :::
Db   9664 GAACAGAAGTCC---CAGAAACCAGAAGTATCTGAGGTCGTAGCTGAGAAGATATCTGAA 9720
```

```
Qy      28 GlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspPro 47
Db    9721 GGAAAGATTGAAGAACCAAAGAAACCTGAAGAAATGGACACTGAAGCTAAATCAGAGAAA 9780

Qy      48 AlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAsnGlyValAlaGlu--- 66
Db    9781 GCTACAGTCTTGGATAAGCAAGTCCTGGAAGAAAAAGAACTAGAGGCTAGTGCAGAAAAG 9840

Qy      67 GlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGlnGluGluGluValValAsp 86
Db    9841 CAAGGTGATCAGGATGTCGAAAAGAAGTCCCAGAAACCAGAAGTATCTGAGGTCGTAGCT 9900

Qy      87 GluAspValGlyGlnArgGluSerGluAspValArgGlu--------------------- 99
Db    9901 GAGAAGATATCTGAAGAAACAATTGAAGAGCCAAAGAAACCTGAAGTAAAGGACACTGAG 9960

Qy     100 ----------------------------------------LysAspArgValGluGlu 105
Db    9961 ATTAAATCAGAGAAAGCCACAGCCTTGGATAAGCAAGTCCTGGAAGAAAAGGAACTAGAG 10020

Qy     106 MetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu---GluThr 124
Db   10021 GCTAGTGCACAAAAGCAATGTGATCAGGATGTCGAAAAGAAGTCCCAGAAACCAGAAGTA 10080

Qy     125 SerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnPro 144
Db   10081 TCTGAGGTCGTAGCTGAG---AAAATATCTGAAGAACCAATTGAAGAGCCAAAGAAACCT 10137

Qy     144 ------------------------------------------------------------ 144
Db   10138 GAAGTAAAGGAAACTGAGGTTAAATCAGAGAAAGCTACAGTCTTGGATAAGCAAGTCCTG 10197

Qy     145 -----------AlaGluSerGlnAlaAsnAspValGly--------------------- 153
Db   10198 GAAGAAAAGGAACTAGAGGCTAGTGCACAAAAGCAAGGTGATCAGGATGTCGAAAAGAAG 10257

Qy     154 PheLysLys------ValPheLysPheValGlyPheLysPheThrValLys--------- 168
Db   10258 TTCCAGAAAGCAGAAGTATCTGAGGTCGTAGCTGAGAAGATATCTGAAGAAACAATTGAA 10317

Qy     169 --------------------LysAsp---------LysAsnGluLysSerAspThrVal 178
Db   10318 GAGCCAAAGAAACCTGAAGTAAAGGACACTGAGATTAAATCAGAGAAAGCCACAGCCTTG 10377

Qy     179 GlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGly 198
Db   10378 GATAAGCAAGTCCTGGAAGAAAAGGAACTAGAGGCTAGTGCACAAAAGCAAGGTGATCAG 10437

Qy     199 Asp-----------HisGlnGluProSerValGluThrAlaValGlyGluSerAlaSer 214
Db   10438 GATGTCGAAAAGAAGTCCCAGAAACCAGAAGTATCTGAAGTCATAGCTGAGAAGATATCT 10497

Qy     215 LysGluSer---GluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGlu 233
Db   10498 GAGGAAAAGATTGAAGAACCAAAGAAACCTGAAGAAAAGGAAACTGAGGTTAAATCAGAG 10557

Qy     234 GlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAla 253
Db   10558 AAA---GCTACAGTCTTGGATAAGCAAGTCCTGGAAGAAAAGGAACTAGAGGCTAGTGCA 10614

Qy     254 LysAspGluGlyGlu---------GluLysGlnGluLysGluProThrLysSerPro--- 269
Db   10615 CAAAAGCAAGGTGATCAGGATGTCGAGAAAAGGTCCCAGAAACCAGAAGTATCTGAGGTC 10674

Qy     270 ----------------------------GluSerProSerSerProValAsnSerGlu 279
Db   10675 GTAGCTGAGAAGGTATCTGAAGGAAAGATTGAAGAACCAAAGAAACCTGAAGTAAAGGAA 10734

Qy     280 ThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGly---TrpArgLysLys 298
Db   10735 ACAGAAGCTAAATCAGAGAAAGCCACAACTTT-GGATATGCAAGTCCTGGAAGAAAGAGA 10793

Qy     299 Thr-SerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGl 318
Db   10794 ACTAGAGGCTAGTGCACAAAAGCAAGGTGATCAGGATGTCGAAAAGAAGTCCCAGAAACC 10853

Qy     318 nGlu---------------AlaGluLysValAspGluGlu-------------------- 326
Db   10854 AGAAGTATCTGAAGTCATAGCTGAGAAGATATCTGAAGAAAAGATTGAAGAACCAAAGAA 10913

Qy     327 ----GluLysGluLysThrGluProAlaSerGlu------------------------- 336
Db   10914 ACCTGAAGAAAAGGAAACTGAGGTTAAATCAGAGAAAGCTACAGTCTTGGATAAGCAAGT 10973

Qy     337 ----GluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGlu-- 354
Db   10974 CCTGGAAGAAAAAGAACTAGAGGCTAGTGCACAAAAGCAAGGTGATCAGGATGTCGAAAA 11033

Qy     355 -----------------------------------------------------LysVa 356
Db   11034 GAAGTCCCAGAAACCAGAAGTATCTGAGGTCGTAGCTGAGAAGGTATCTGAAGGAAAGAT 11093

Qy     356 lGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAl 376
Db   11094 TGAAGAACCAAAGAAACCTGAAGTAAAGGAAACTGAGGTTAAATCA---GAGAAAGCCAC 11150

Qy     376 aProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGl 396
Db   11151 AACTTTGGATAAGCAAGTCCTGGAAGAAAAA----------------------------- 11181

Qy     396 uValHisValSerThrValGluLysThrGluGluGluGlnGlyGly---GlyGlyGluAl 415
Db   11182 -------------GAACTAGAGGCTAGTGCACAAAAGCAAGGTGATCAGGATGGAAAATC 11228

Qy     415 aGluGlyGlyValVal-----------------------------ValGluGlyThrGl 425
Db   11229 TCGAGATGATATTATTAAAACATTGAAAGAAAGGCTCACAGAATTATCCAAGGCATTAGG 11288

Qy     425 yGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluPr 445
Db   11289 ATCTAGTGTC---GATGAAATTTTGAGAGAAAGTCGGGAAATTGTCAATAACCTTGAAGA 11345

Qy     445 o-----------AlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAs 461
Db   11346 TGATAAGGTAGTGGCAAAACATTTATTCAAATTAAGAGACCATATAGTTCACACCTATGA 11405

Qy     461 pHisThrGln-------------------------------------------------Le 465
Db   11406 CGGAAAAAGAGGAGAAGAAAATAAAGAAAAAGAACTCTTCGAATCCTTTATTGAATTGTT 11465

Qy     465 uThrAspLeuSerProGlu----------------------------------------- 471
Db   11466 GTGTGAGGCCTCCCCTGAAGCAGCTGAGAAAGTAAAACTCAACTATCTTAAGGAAATTAA 11525

Qy     471 ------------------------------------------------------------ 471
Db   11526 AACCAATGTGATTTTAACAAAAGCTACTATTCAACTTATCGACGACTCTAACATGTTCAC 11585

Qy     471 ------------------------------------------------------------ 471
Db   11586 CAAACCGTCCTTATTAATCCCCAAACTTTTAAATTTGGAGAGAGTTGCCGTTAAAATTCA 11645

Qy     472 ----GluLysThrLeuProLysHisProGluGlyIleValSer----------------- 484
Db   11646 ATCAGAAACATATGTTGATAAGTCATCGGAAAAAATGATTAGTTTGCAACAGAGTCTTAT 11705

Qy     484 ------------------------------------------------------------ 484
Db   11706 GGATATTTTCGTAATTCTTGATGACTTTTTAGATGACGAGACGGAAGTTTTAAAACCAAA 11765

Qy     485 ------------GluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySe 500
Db   11766 AATTGAAAATATTAAAACAACGTTGCTCAGTGACTATGATTATATTGAAAAAAAAGATGG 11825

Qy     500 rProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLy 520
```

```
Db  11826 ACCCCTCTTAACGGCTGTAATAAATGGAAAGATTAATGTTGTATCC-------------- 11871
Qy    520 sGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThr-- 539
Db  11872 ------------------------------------------CAACATATCTTAACTAT 11888
Qy    540 ---------------GluSerProGluSerAlaAspGlu---GlnLysGlyGluSerSe 553
Db  11889 TATTGAAGAAGTTAACCAATTGACTGAAAATCATGATCAAAAGGAAAAAGACGTGTCCAA 11948
Qy    553 rAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAl 573
Db  11949 CGCGGAGGCGGATAATTTCGCCGACGAAAAGCGA------GAAGAAAGTCAAAAGGAGGA 12002
Qy    573 aProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGl 593
Db  12003 GATTAAGGACTCTGAAGCAAAGCACAAGAAATCGAAGGTTTCGGAGAAGAAATCA----- 12057
Qy    593 yIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArgPr 613
Db  12058 -------------ATCGAAGAAGAGAAACTTGAAGACAAGAAAGAAAAACAGACAGAGTC 12104
Qy    613 oSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSerTh 633
Db  12105 TGCTATTGATGAAAAATCCCAAAAGGCTGAGGTGTCCGAAACTGTCTCCGAAAAGATTAC 12164
Qy    633 rAspSerThrValSerGluMetGln---AspGluValLysThrValGlyGluGluGlnLy 652
Db  12165 TGATGAGAAGGCACAAGAAAGTCAAAAGGAGGAGGTTAAG------GATTCTGAAGCCAA 12218
Qy    652 sProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVa 672
Db  12219 GCCTAAGAAGGCAAAGGTCTTGGAGAAGAAATCAATCGAAGAAGAGAAACTT-------- 12270
Qy    672 lGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProAr 692
Db  12271 -------GAAGACAAGAAAGAAAAACAGACGGAGTCTGCTATTGATGAA----------- 12312
Qy    692 gThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyTh 712
Db  12313 ------------AAATCCCAAAAGGCTGAGGTGTCCGAAATTGTCTCCGAAAAGATTAC 12359
Qy    712 rAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGl 732
Db  12360 TGATGAGAAGGCACAAGAAAGCCAAAAGAAGGAGGTTAAGGATTCTGAAGCCAAGCCTAA 12419
Qy    732 uProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLe 752
Db  12420 GAAGGCA-----------------------------------------------AAGGT 12431
Qy    752 uValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGlu---------As 769
Db  12432 CTTGGAGAAGAAATCAATCGAAGAAGAGAAACTTGAAGACAAGAAAGAAACACAGACGGA 12491
Qy    769 pSerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSe 789
Db  12492 CTCTGCTATTGATGAAAAATCCCAAAAGGCTGAGGTGTCCGAAATTGTCTCC-------- 12543
Qy    789 rIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAl 809
Db  12544 ---------------------------GAAAAGATTACTGATGAGAAGGCACAAGAAAG 12575
Qy    809 aThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaVa 829
Db  12576 TCAAAAGGAGGAG---------GTTAAGGATTCTGAAGCCAAGCCTAAGAAGGCAAAGGT 12626
Qy    829 lValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnTh 849
Db  12627 CTTG---------GAGAAGAAATCAATCGAAGAAGAGAAACTTGAAGACAAGAAAGAAAA 12677
Qy    849 rGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer-LysThrLeuV 869
Db  12678 ACAGACGGAGTCTGCTATT-----------------GATGAAAAATCCCAAAAGGCTG- 12718

Qy    869 alHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluA 889
Db  12719 -----------------------------------------------------------A 12719
Qy    889 rgSerProSerTrpIleSerAlaSerValThr--------------------GluProL 902
Db  12720 GGTGTCCTGAAATTGTCTCCGAAAAGATTACTGATGAGAAGGCACAAGAAAGCCAAAAGA 12779
Qy    902 euGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleI 922
Db  12780 AGGAGGTTAAGGGTTCTGAAGCCAAGCCTAAGAAGGCAAAGGTCTTGGAGAAGAAATCAA 12839
Qy    922 leAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspA 942
Db  12840 TCGAGGAAGAG-----------------AAACTTGAAAACAAGAAAGAAACACAGACGG 12881
Qy    942 spMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerG 962
Db  12882 AC---------TCTGCTATTGAT--------------------------GAAAAATCCC 12905
Qy    962 lu-AlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMet 981
Db  12906 AAAAGGCTTGAGGTGTCCGAAATTGTC------------TCCGAAAAGATTACTGATGAG 12953
Qy    982 ValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProVal 1001
Db  12954 AAGGCACAAGAAAGTCAAAAGGAGGAGGTTAAGGATTCTGAAGCCAAGCCTAAGAAAGCA 13013
Qy   1002 GlnGluValGluSerGlyValLeuAspThrGlu------GluGluGluArgGlnThrGln 1019
Db  13014 AAGGTCTTGGAGAAGAAATCAATCGAAGAAGAGAAACTTGAAGACAAGAAAGAAACACAG 13073
Qy   1020 AlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGln 1039
Db  13074 ACG------GACTCTGCTATTGATAAAAATCCCAAAAGGCTGAGGTGTCCGAAATTGTCT 13127
Qy   1040 ThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSer 1059
Db  13128 CCGAAAAGATTACTTGATGAGAAGGCACAAGAAAGTCAAAAGGAGGAGGTTAAGGATTCT 13187
Qy   1060 GluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAla 1079
Db  13188 GAA------GCCAAGCCTAAGAAAGCAAAGGTCTTGGAGAAGAAATCAATCGAAGAAGCG 13241
Qy   1080 GlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGlu 1099
Db  13242 AAACTTGAAGACAAGAAAGAA---ACACAGACGGAGTCTGCTATTGATGAAAAATCCCAA 13298
Qy   1100 ValProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGln 1119
Db  13299 AAGGCTGAGGTGTCCGAAATTGTCTCCGAAAAGATTACTGATGAG------AAGGCACAA 13352
Qy   1120 GlnLeuMetGluGlnAlaVal---------------AlaProGluSerSerGluThrLeu 1134
Db  13353 GAAAGCCAAAAGAAGGAGGTTAAGGGTTCTTAAGCCAAGCCTAAGAAAGCAAAGGTCTTG 13412
Qy   1135 ThrAspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThr 1154
Db  13413 ---------GAGAA-AGAAATCAATCGAAGAAGAGAAACTTGAAGACCAAGAAAGAAAAA 13462
Qy   1155 GlnGlnAspGluThrIleAspSerGlnAspSerLysAla----ThrAlaAlaValArgGl 1173
Db  13463 CAGACGGAGTCTGCTATTGATGAAAAATCCCAAAAGGCTTGAGGTGTCCGAAATTGTCTC 13522
Qy   1173 nSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGlu---------ProSe 1190
Db  13523 CGAAAAGATTACTGATGAGAAGGCACAAGAAAGCCAAAAGAAGGAGGTTAAGGATTCTGA 13582
Qy   1190 rThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspVa 1210
Db  13583 AGCCAAGCCTAAGAAGGCAAAGGTCTTGGAGAAGAAATCAATCGAA------GAAGAGAA 13636
```

```
Qy     1210 lLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGl 1230
              ||||||     :::::::::   |||    :::|||:::                ||
Db    13637 ACTTGAAAACAAGAAAGAAAAACAGACGGAGTCTGCTATT---------------GATGA 13681

Qy     1230 uValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGluGluGlnGl 1250
            |     |||:::   ||||||                ||||||:::   :::|||:::
Db    13682 AAAATCCCAAAAGGCTGAGGTGTCCGAAACTGTCTCCGAAAAGATTACTGATGAGAAGGC 13741

Qy     1250 uValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGl 1270
                                            :::       :::      ||||||||
Db    13742 A-----------------------GAAGAAAGCCGAAAGGAGGAGGTTAAGGACTCTGA 13777

Qy     1270 uValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSerLeuSerLe 1290
            |                 :::    |||::::::     ::::::     |||::
Db    13778 A------------------GCCAAGCCTAAGAAAGCAAAGGTCTTGGAGAAGAAATCAAT 13819

Qy     1290 uGluGluGlyGluMetGluThrAspValGluLys-------------------------- 1301
            :||||||   ::::::|||        ||||||
Db    13820 CGAAGAAGAGAAACTTGAAGACAAGAAAGAAAAACAGACAGAGTCTGCTATTGATGAAAA 13879

Qy     1302 ------------GluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGl 1317
                        |||   ||||||    |||::::::::::::|||      ||
Db    13880 ATCCCAAAAGGCTGAGGTGTCCGAAACTGTCTCCGAAAAGATTACTGATGAGAAGGCACA 13939

Qy     1317 nGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMe 1337
            ||||:::        |||   :::   :::      ||||||
Db    13940 AGAAAGCCAAAAGAAGGAGGTTAAGGATTCTGAAGCCAAGCCT----------------- 13982

Qy     1337 tProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAs 1357

Db    13982 ------------------------------------------------------------ 13982

Qy     1357 pGlnAspLysAlaGlyCysIleGlu---------ValGlnValGlnSerLeuAspThrTh 1374
                   ||||||     :::|||      ::::::::::::   ||||||
Db    13983 ----AAGAAAGCAAAGATTTTGGAGAAGAAATCAATCGAAATAGAGAAATTGGAT----- 14033

Qy     1374 rValThrGlnThrAlaGluAlaValGluLysValIleGluThrValValIleSerGluTh 1394
                              |||   ||||||    ||||||   |||   ::::::||
Db    14034 ----------------GAAAAGAAAGAAAAGCAGACGGAGACTAAGGTTGCCACTGATAC 14077

Qy     1394 rGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaTh 1414
            |         :::   |||         :::      ||||||   |||
Db    14078 C------AAATCCCAAACGGTTGAGGTTTCCGAAATCGTTTTGGAAAAGATATCGGAA-- 14129

Qy     1414 rGlyGlyHisTrpThrLeuGlnHisAlaGluAspThr------ValProLeuGlyProGl 1432
                              :::   ||||||::::::      |||   |||      ::
Db    14130 ------------------GAAAAGGCAGAAGAAAGCCAAAAGGTTGAACTT---AAGGA 14167

Qy     1432 uSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisPr 1452
            :|||:::|||:::|||               |||            :::|||||||||
Db    14168 CTCTGAGGCCAAATCAAAGAAGGCAAAGGTCCTG---GAAAAGAAGTCAACCTTGAAAGA 14224

Qy     1452 oAspLeuGlnGlyGluIleSerAlaSerGlnArgGlu---------------ArgSerGl 1467
               |||       |||      |||:::|||                  :::|||::
Db    14225 GAAACTCGAT---GAGAATGATAAAAAACAGAAGGAGGATGGCGCTACTAACAAATCCCA 14281

Qy     1467 uGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAla----- 1485
            ::::   :::      |||      |||:::         :::|||      |||
Db    14282 AAAGGCAGAGGCTGCTGATGTTGTTCCCGAAAAAATATCTGAAGAGAAGGTTGCTGAAAT 14341

Qy     1486 ----------------IleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGl 1500
                            ::::::         |||::::::|||:::   |||      ::
Db    14342 AAAAACTCCTGAACCTATGGACTCAAAGGCTAAATCTAAACCAGATGGATTACCTGCCGA 14401

Qy     1500 uSerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaAr 1520
            :    ||||||:::          ::::::                ::::::      ::
Db    14402 TGAAAAGAGTCACGGGGCTAAGGTTTCTGAATCTGTTCCTGTAAAGAATGAAGCTGAGAA 14461

Qy     1520 gThrGlu------ThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAl 1538
            :|||:::        :::|||   :::            |||         ||||||
Db    14462 AACAGATCAACTTTCAGCAAAGAAACCAACTGTTCTGGATGAGGATCTGGTTGTTCCAAA 14521

Qy     1538 aCysArg----------------------------------------------------- 1540
                   :::
Db    14522 ACGAAAGCCATATTTGGCTGAACAAACTGCAGATTCTATTAGCCTGCAAACGTACAAGTC 14581

Qy     1541 -LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHi 1560
              :::||||||          :::|||         :::   :::   ||||||   |||
Db    14582 CATGGACTCCGAGTACAAGGACCGA------AAGGAATCCCGAAGTGCCAAGCGGAAG-- 14633

Qy     1560 sProValProGlnProArgGluAspLeuGlnVal 1571
                              |||   |||:::|||:::
Db    14634 ------------CCCACGGTTGACATCCAACTA 14654
```

RESULT 15

```
AAS80823
ID   AAS80823 standard; cDNA; 7568 BP.
XX
AC   AAS80823;
XX
DT   13-FEB-2002  (first entry)
XX
DE   DNA encoding novel human diagnostic protein #16627.
XX
KW   Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW   food supplement; medical imaging; diagnostic; genetic disorder; ss.
XX
OS   Homo sapiens.
XX
PN   WO200175067-A2.
XX
PD   11-OCT-2001.
XX
PF   30-MAR-2001; 2001WO-US08631.
XX
PR   31-MAR-2000; 2000US-0540217.
PR   23-AUG-2000; 2000US-0649167.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Drmanac RT,  Liu C,  Tang YT;
XX
DR   WPI; 2001-639362/73.
DR   P-PSDB; ABG16636.
XX
PT   New isolated polynucleotide and encoded polypeptides, useful in
PT   diagnostics, forensics, gene mapping, identification of mutations
PT   responsible for genetic disorders or other traits and to assess
PT   biodiversity  -
XX
PS   Claim 1; SEQ ID No 16627; 103pp; English.
XX
CC   The invention relates to isolated polynucleotide (I) and
CC   polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC   polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC   and gene mapping, and in recombinant production of (II). The
CC   polynucleotides are also used in diagnostics as expressed sequence tags
CC   for identifying expressed genes. (I) is useful in gene therapy techniques
CC   to restore normal activity of (II) or to treat disease states involving
CC   (II). (II) is useful for generating antibodies against it, detecting or
CC   quantitating a polypeptide in tissue, as molecular weight markers and as
CC   a food supplement. (II) and its binding partners are useful in medical
CC   imaging of sites expressing (II). (I) and (II) are useful for treating
CC   disorders involving aberrant protein expression or biological activity.
CC   The polypeptide and polynucleotide sequences have applications in
CC   diagnostics, forensics, gene mapping, identification of mutations
CC   responsible for genetic disorders or other traits to assess biodiversity
CC   and to produce other types of data and products dependent on DNA and
CC   amino acid sequences. AAS64197-AAS94564 represent novel human
CC   diagnostic coding sequences of the invention.
CC   Note: The sequence data for this patent did not appear in the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 7568 BP; 2332 A; 1873 C; 1825 G; 1538 T; 0 other;
```

```
Alignment Scores:
Pred. No.:              3.28e-16      Length:          7568
Score:                  492.00        Matches:         386
Percent Similarity:     33.47%        Conservative:    269
Best Local Similarity:  19.72%        Mismatches:      714
Query Match:            6.09%         Indels:          589
DB:                     23            Gaps:            75

US-09-902-432-4 (1-1596) x AAS80823 (1-7568)

Qy      45 AlaAspProAlaThrLysLeu---------------ProGlnLysAsnGlyGlnLeuSer 59
Db    1593 GCAAACCCTGCGGAGAAAATCATCCGAGTCCTGTTTCCTGGGAACAGCACCCAGTACAAC 1652

Qy      60 SerValAsnGlyValAlaGluGlnGlyAspValHis------------------ValGln 73
Db    1653 ATCCTGGAAGGGTTGGAAAAGCTCAAACATCTAGACTTTCTGAAGCAGCCACTGGCCACC 1712

Qy      74 GluGluAsnGlnGluGlyGlnGluGluGluValValAspGluAspValGlyGlnArgGlu 93
Db    1713 CAAAAGGATCTCACTGGCCAGGTGCCCACTCCTGTGGTGAAACAAACAAAACTGAAACAG 1772

Qy      94 SerGluAspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaVal 113
Db    1773 AGGGCTGATAGCCGAGAAAGT---------CTGAAGCCAGCCGCAAAACCACTTCCTAGC 1823

Qy     114 GluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIleGluGlnIleProAla 133
Db    1824 AAATCCGTGCGCAAGGAGTCAAAAGAAGAAACCCCTGAGGTC---------------ACA 1868

Qy     134 SerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGly 153
Db    1869 AAAGTGAATCACGTGGAA---------AAGCCACCCAAAGTTGAAAGCAAAGAA------ 1913

Qy     154 PheLysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGlu 173
Db    1914 ------------------------------AAGGTAATGGTGAAAAAAGACAAGCCAGTA 1943

Qy     174 LysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAla 193
Db    1944 AAAACAGAGACCAAACCTTCAGTGACTGAAAAGGAG------------------------ 1979

Qy     194 SerValGlyAlaGlyAspHisGlnGluProSer---ValGluThrAlaValGlyGluSer 212
Db    1980 ---------GTTCCCAGCAAAGAAGAGCCATCTCCAGTGAAAGCCGAGGTGGCTGAG--- 2027

Qy     213 AlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGln 232
Db    2028 ---------------------AAGCAAGCCACAGATGTCAAACCCAAAGCTGCCAAGGAG 2066

Qy     233 GluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGlu 252
Db    2067 AAGACGGTGAAAAAGGAAACAAAGGTAAAGCCTGAAGACAAGAAAGAGGAGAAAGAAAAG 2126

Qy     253 AlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerPro 272
Db    2127 CCAAAGAAAGAAGTGGCTAAAAAGGAGGACAAAACACCTATCAAGAAGGAGGAAAAACCA 2186

Qy     273 SerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrp 292
Db    2187 AAA---------AAGGAAGAGGTGAAAAAAGAAGTCAAAAAAGAGATC------------ 2225

Qy     293 AlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGluThrAla 312
Db    2226 ---------AAGAAAGAAGAGAAAAAAGAACCCAAGAAAGAGGTTAAGAAAGAAACACCG 2276

Qy     313 GluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGluLysThrGlu 332
Db    2277 CCAAAGGAAGTCAAGAAGGAAGTTAAGAAG---------------GAAGAGAAGAAGGAA 2321

Qy     333 ProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAsp 352
Db    2322 GTGAAAAAGGAAGAAAAGGAACCCAAAAAAGAAATT------AAGAAGCTCCCTAAAGAC 2375

Qy     353 TyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGlu 372
Db    2376 GCAAAGAAATCATCTACTCCTCTG---------------------------TCTGAAGCA 2408

Qy     373 GluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGlu 392
Db    2409 AAAAAACCAGCTGCTTTAAAACCAAAAGTACCCAAGAAGGAAGAGTCTGTCAAGAAAGAT 2468

Qy     393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db    2469 TCTGTTGCTGCCGGAAAGCCAAAGGAGAAGGGGAAAATAAAAGTCATTAAGAAGGAAGGC 2528

Qy     413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db    2529 AAGGCCGCAGAGGCTGTCGCTGCAGCTGTCGGCACTGGAGCCACCACAGCAGCTGTCATG 2588

Qy     433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db    2589 GCGGCAGCTGGAATAGCAGCC---------ATTGGCCCTGCCAAAGAACTCGAAGCTGAG 2639

Qy     453 ArgGluMetCysValSerGlyGlyAspHisThrGln---LeuThrAspLeuSerProGlu 471
Db    2640 AGGTCCCTTATGTCATCTCCTGAGGATCTAACCAAGGACTTTGAAGAGTTAAAGGCTGAA 2699

Qy     472 GluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSer 491
Db    2700 GAGGTCGATGTAACAAAG---------GACATCAAGCCTCAGCTGGAGCTAATCGAAGAC 2750

Qy     492 GlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeu 511
Db    2751 GAAGAGAAACTGAAG---GAAACTGAGCCAGTCGAAGCCTACGTCATCCAGAAGGAGAGA 2807

Qy     512 LysLysLeuSerGlyLysLysGlnLysGlyLysArgGly--------------------- 524
Db    2808 GAAGTCACCAAAGGTCCTGCCGAGTCCCCTGATGAGGGAATCACTACCACTGAAGGGGAG 2867

Qy     525 GlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSer 544
Db    2868 GGCGAATGTGAACAGACACCTGAGGAGCTGGAGCCCGTCGAGAAGCAGGGAGTAGAC--- 2924

Qy     545 AlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThr 564
Db    2925 ---GACATTGAAAAATTTGAAGATGAAGGAGCCGGTTTTGAAGAATCTTCAGAGACT--- 2978

Qy     565 CysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThr 584
Db    2979 ------------GGAGACTATGAAGAGAAGGCAGAAACTGAGGAGGCTGAGGAGCCAGAA 3026

Qy     585 SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetVal 604
Db    3027 GAGGATGGGGAGGAACAC---------------------------------------GTA 3047

Qy     605 ThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeuGlu--- 623
Db    3048 TGTGTGAGCGCCTCCAAGCACAGCCCCACTGAGGATGAGGAAAGTGCCAAGGCGGAGGCT 3107

Qy     624 ------------LysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGlu 639
Db    3108 GATGCATACATCAGGGAGAAGAGGGAGTCTGTGGCCAGTGGGGATGACCGAGCCGAAGAA 3167

Qy     640 MetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArg 659
Db    3168 GACATGGATGAGGCCATTGAGAAAGGAGAGGCTGAACAATCTGAAGAGGAGGCTGATGAG 3227

Qy     660 ValAspThrSer------------------------------ValSerTrpGluAlaLeu 669
Db    3228 GAGGACAAAGCTGAAGATGCCAGAGAGGAGGAATATGAGCCGGAAAAAATGGAAGCTGAA 3287

Qy     670 IleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGly 689
Db    3288 GACTATGTGATGGCTGTGGTCGACAAGGCTGCAGAGGCTGGTGGTGCCGAGGAGCAGTAT 3347
```

```
Qy    690 Gly---------------ProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAla 704
Db   3348 GGATTCCTCACCACACCAACCAAGCAACTAGGAGCCCAGTCTCCTGGCCGAGAACCTGCA 3407

Qy    705 SerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGln 724
Db   3408 TCTTCAATTCATGAT---------GAGACTTTACCTGGAGGCTCAGAGAGCGAGGCCACC 3458

Qy    725 AlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSer 744
Db   3459 GCTTCTGATGAGGAGAATCGAGAA------GACCAGCCTGAGGAATTCACTGCCACCTCT 3512

Qy    745 ThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGlu 764
Db   3513 GGCTAC------------------------------------ACTCAGTCTACTATTGAG 3536

Qy    765 GluLysAlaGluAspSerSerValGluGlnLeuSerThr--------------------- 777
Db   3537 ATATCCAGTGAGCCCACCCCCATGGATGAGATGTCTACCCCTCGAGACGTGATGAGTGAT 3596

Qy    778 ------------GluIleGluProSerArgGluGluSerTrpValSerIleLysLys--- 792
Db   3597 GAGACCAACAATGAAGAGACGGAGTCCCCTTCTCAGGAATTCGTAAATATCACCAAATAT 3656

Qy    792 ------------------------------------------------------------ 792
Db   3657 GAATCTTCATTGTATTCTCAGGAATACTCTAAACCTGCTGATGTTACACCGCTCAACGGA 3716

Qy    793 PheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGlu 812
Db   3717 TTTTCTGAAGGATCAAAAACAGATGCCACTGATGGCAAGGATTACAATGCTTCAGCCTCT 3776

Qy    813 AspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValValProLeu 832
Db   3777 ACCATATCACCACCC--------------------------------------------- 3791

Qy    833 SerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGluLeuPro 852
Db   3792 ---------TCTTCCATGGAGGAAGACAAATTCAGCAGATCTGCT--------------- 3827

Qy    853 GlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrVal 872
Db   3828 ------TTACGTGATGCTTACTGCTCTGAAGTGAAAGCCAGCACC--------------- 3866

Qy    873 SerValAlaValIleAspGlyThrArgAlaValThrSerValGlu---------GluArg 889
Db   3867 ACTTTGGACATCAAAGATAGCATCTCAGCTGTTTCAAGTGAAAAGGTCAGCCCATCGAAG 3926

Qy    890 SerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr---AlaGlyGlu 908
Db   3927 AGCCCGTCC---CTGAGTCCATCTCCACCATCACCCTTAGAAAAGACCCCCCTGGGTGAA 3983

Qy    909 ---------------AlaMetProProValGlu---GluValThrGluLysAspIleIle 922
Db   3984 CGTAGTGTGAACTTCTCTCTGACGCCCAATGAGATTAAAGTCTCTGCAGAGGCAGAAGTA 4043

Qy    923 AlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAsp 942
Db   4044 GCCCCGGTGTCTCCTGAGGTGACCCAAGAAGTAGTTGAAGAACATTGTGCTAGTCCTGAG 4103

Qy    943 MetValThrSerGluValAspPheThrSerGluAlaValThrAlaThr------------ 958
Db   4104 GACAAGACTCTGGAAGTGGTGTCACCATCTCAGTCCGTGACTGGCAGTGCTGGTCACACA 4163

Qy    959 ------------------------GluThrSerGluAlaLeuArgThrGluGluVal--- 969
Db   4164 CCTTACTATCAATCTCCTACTGACGAGAAATCCAGTCATCTCCCTACAGAAGTCATTGAA 4223

Qy    969 ------------------------------------------------------------ 969
Db   4224 AAACCACCAGCAGTTCCAGTGAGTTTTGAATTCAGTGATGCCAAAGATGAGAATGAAAGG 4283

Qy    969 ------------------------------------------------------------ 969
Db   4284 GCTTCAGTAAGCCCCATGGATGAGCCCGTGCCTGACTCAGAGTCTCCTATTGAAAAAGTT 4343

Qy    969 ------------------------------------------------------------ 969
Db   4344 TTGTCTCCTTTACGCAGCCCGCCCCTCATTGGATCCGAGTCTGCTTATGAAAGTTTTCTA 4403

Qy    970 ---------------------------------------------ThrGluAlaSerGly 974
Db   4404 AGTGCTGATGACAAGGCTTCTGGCAGAGGTGCCGAAAGTCCTTTTGAAGAAAAGAGTGGA 4463

Qy    975 AlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSer--------- 991
Db   4464 AAACAAGGCTCTCCAGACCAAGTAAGTCCAGTTTCTGAAATGACTTCTACTAGTCTTTAC 4523

Qy    992 ProAspThrThrGlu------------GluAlaThrProValGlnGluValGluSerGly 1007
Db   4524 CAAGACAAACAGGAAGGGAAAAGCACAGACTTTGCACCAATAAAAGAAGACTTTGGCCAA 4583

Qy   1008 ValLeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAsp 1027
Db   4584 GAAAAGAAAACTGATGATGTTGAAGCCATGAGTTCTCAACCAGCACTGGCTCTGGATGAA 4643

Qy   1028 Lys--------------------------------------------------------- 1028
Db   4644 AGGAAATTAGGAGATGTTTCTCCCACACAAATAGATGTCAGTCAGTTTGGATCTTTTAAA 4703

Qy   1029 ---------------------------------ValLysGluGluSerGlnValProAla 1037
Db   4704 GAAGACACTAAGATGTCCATTTCTGAAGGTACTGTCTCAGACAAGTCAGCTACTCCTGTT 4763

Qy   1038 ThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   4764 GATGAGGGCGTAGCAGAAGACACGTACTCTCATATGGAGGGTGTGGCCTCAGTGTCCACA 4823

Qy   1058 AspSerGluValLeuAlaSerGluLysGlu---------LysAspValMetProLysGly 1074
Db   4824 GCCTCAGTGGCTACGAGCTCATTTCCAGAGCCAACAACAGATGATGTGTCTCCATCTCTG 4883

Qy   1075 ProValGlnGluAlaGlyAlaGluHis--------------------------------- 1083
Db   4884 CATGCT---GAGGTTGGCTCCCCACATTCCACAGAAGTAGATGACTCCCTTTCAGTGTCT 4940

Qy   1084 LeuAlaGlnGlySerGluThrGlyGlnAlaThr--------------------------- 1094
Db   4941 GTTGTGCAAACACCTACCACATTCCAGGAAACAGAAATGTCTCCATCTAAAGAAGAATGC 5000

Qy   1095 ------ProGluSerLeuGluValProGluVal------------------------Thr 1104
Db   5001 CCAAGACCGATGTCAATTTCTCCACCAGATTTCTCCCCTAAAACTGCAAAGTCCAGGACA 5060

Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeu---------GlnGlnLeu 1121
Db   5061 CCCGTTCAAGATCACAGATCTGAACAGTCCTCAATGTCTATTGAATTTGGCCAAGAATCT 5120

Qy   1122 MetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThr------ 1139
Db   5121 CCTGAGCAATCCCTTGCTATGGACTTCAGTCGACAGTCTCCAGATCACCCTACAGTGGGT 5180

Qy   1140 ------------------------AsnGlySerThrProLeu------AlaAspSerAsp 1149
Db   5181 GCAGGCGTGCTTCACATCACTGAAAATGGGCCAACTGAAGTGGACTACAGTCCTTCTGAC 5240

Qy   1150 ThrAlaAspGlyThr------------------GlnGlnAspGluThrIleAspSerGln 1163
Db   5241 ATGCAGGACTCCAGTTTATCACATAAGATACCACCTATGGAGGAGCCGTCCTACACCCAA 5300

Qy   1164 AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThr 1183
Db   5301 GATAATGATCTTTCTGAGCTCATCTCAGTATCTCAGGTA---GAGGCCTCCCCGTCCACC 5357

Qy   1184 AlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGly 1203
```

```
Db   5358 TCTTCTGCTCATACCCCTTCTCAGATCGCTTCT------CCTCTCCAA------------ 5399

Qy   1204 GluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaVal 1223
          |||:::          ||||||      |||          |||      |||:::
Db   5400 GAAGATACTCTATCCGATGTTGCTCCTCCCAGAGATATGTCCTTATATGCCTCACTCACC 5459

Qy   1224 ProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGlu 1243
                                                   :::|||      |||:::||||||
Db   5460 -----------------------------TCTGAAAAAGTGCAAAGTCTGGAAGGAGAG 5489

Qy   1244 LysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAla 1263
          |||:::      :::   ::::::          :::           |||
Db   5490 AAGCTCTCTCCAAAATCTGATATCTCTCCACTCACCCCACGAGAGTCCTCTCCTTTATAT 5549

Qy   1264 AspValThrTyrAspSerGluValMetGlyVal------------AlaGlyCysGlnGlu 1279
                |||:::                      |||          |||   |||
Db   5550 TCACCTACTTTTTCAGATTCTACCTCTGCAGTCAAAGAGAAAACAGCAACTTGCCACAGT 5609

Qy   1280 LysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspVal 1299
                |||         :::   :::   |||   |||                     |||
Db   5610 TCCTCTTCTCCACCAATAGATGCAGCATCCGCAGAGCCCTATGGCTTCCGTGCCTCAGTG 5669

Qy   1299 ------------------------------------------------------------ 1299

Db   5670 TTATTCGATACAATGCAACACCATCTAGCCTTGAATAGAGATTTGTCCACACCTGGCCTG 5729

Qy   1300 GluLysGluLysArgGluThrLysProGluGlnValSer---------GluGluGlyGlu 1316
          ||||||:::                   |||         |||         ::::::   |||
Db   5730 GAGAAGGACAGTGGAGGGAAGACACCTGGTGACTTTAGCTATGCCTATCAAAAGCCTGAG 5789

Qy   1317 GlnGluThrAlaAlaProGluHisGluGly------------ThrTyrGlyLysProVal 1332
          :::   |||   :::|||:::   |||                  :::|||   |||
Db   5790 GAAACAACCAGGTCCCCAGATGAAGAAGATTATGACTATGAGTCTTATGAGAAGACCACC 5849

Qy   1333 LeuThrLeuAspMet--------------------------------------ProSer 1339
                |||   |||:::
Db   5850 CGGACCTCAGATGTGGGTGGCTATTACTATGAGAAGATAGAGAGAACCACAAAATCTCCA 5909

Qy   1340 SerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAsp 1359
          |||:::      |||      :::         ::::::|||      :::         |||:::
Db   5910 AGTGACAGTGGCTACTCCTATGAGACCATTGGGAAAACTACCAAGACCCCTGAAGATGGT 5969

Qy   1360 LysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAla 1379
                                                                     :::
Db   5970 -------------------------------------------------GACTATTCCTAT 5981

Qy   1380 GluAlaValGluLysValIleGluThrVal-----------------------ValIle 1391
          |||   :::||||||            |||                        :::
Db   5982 GAAATTATTGAGAAGACCACACGGACCCCTGAAGAGGGCGGGTACTCATATGACATAAGT 6041

Qy   1392 SerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLysSer 1411
             :::|||         ||||||      |||               |||:::|||
Db   6042 GAAAGACCACCAGCCCCCCCGAAGTGAGTGGTTACAGCTATGAAAAGACTGAGAGGTCT 6101

Qy   1412 ---------------------------------------SerAlaThrGlyGlyHisTrp 1418
                                                  |||      ||||||||
Db   6102 AGAAGGCTTCTGGATGACATCAGCAATGGCTATGATGACTCTGAGGATGGTGGCCAC--- 6158

Qy   1419 ThrLeu-----------------GlnHisAlaGluAspThrValProLeuGlyProGlu 1432
          ||||||                                 :::   :::            ||||||
Db   6159 ACACTTGGGGACCCCAGCTACTCTTATGAAACCACTGAGAAAATTACCAGTTTCCCTGAG 6218

Qy   1433 SerGlnAlaGluSerIleProIleIle-------------------------------- 1441
          |||:::      |||
Db   6219 TCTGAAGGTTATTCCTATGAGACATCTACAAAGACAACACGAACCCCTGATACTTCCACA 6278

Qy   1442 -----------------------ValThrProAlaProGluSerThrLeuHisProAsp 1453
                                 :::|||      |||:::::::::   :::
Db   6279 TACTGTTACGAGACTGCAGAGAAAATCACTAGAACCCCTCAGGCATCCACATATTCCTAC 6338

Qy   1454 LeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysProAsp 1473
                ::::::         :::      ::::::|||         |||   :::   |||
Db   6339 GAGACTTCAGACCTATGCTACACTGCAGAAAAGAAGTCCCCCTCAGAAGCCCGTCAGGAT 6398

Qy   1474 AlaGly-------------------------------ProAspAlaAspGlyLysGlu 1482
                                                |||      :::
Db   6399 GTCGATTTATGCCTCGTGTCCTCTTGTGAATACAAGCACCCCAAGACAGAGCTTTCACCC 6458

Qy   1483 Ser-ThrAlaIleGluLysValLeu---------------------------------- 1490
          |||   :::   |||      :::|||
Db   6459 TCTTTCATTAATCCCAATCCTCTTGAGTGGTTTGCCAGTGAAGAACCCACTGAAGAATCT 6518

Qy   1491 ------------------LysAlaGluProGluIleLeuGluLeuGluSerLysSerAs 1504
                            :::   ||||||      |||:::   ||||||
Db   6519 GAAAAGCCCCTCACTCAATCAGGGGGAGCCCCACCGCCTCCAGGAGGAAAGCAACAGGGC 6578

Qy   1504 nLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThr-- 1523
                |||:::      :::   |||:::|||      :::      |||      ::::::|||
Db   6579 CGACAGTGTGATGAAACCCCTCCC-ACCTCAGTCAGCGAGTCAGCCCCATCCCAGACCGA 6637

Qy   1524 ----------AlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysAr 1540
                    |||||||||                  |||   |||
Db   6638 CTCTGATGTTCCCCCGGAGACTGAAGAGTGCCCCTCCATCACG------GCCGATGCCAA 6691

Qy   1540 gLeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHi 1560
           :::||||||   :::   :::                  |||            |||||
Db   6692 TATCGACTCTGAAGACGAGTCGGAAACCATCCCCACAGACAAAACTGTCACGTACAAACA 6751

Qy   1560 s-----------------ProVal------------ProGlnProArgGlu-------- 1567
          |                 ||||||               |||   ||||||
Db   6752 CATGGACCCACCTCCAGCTCCCGTGCAAGACCGCAGCCCTTCGCCACGCCACCCTGATGT 6811

Qy   1568 ----------AspLeuGlnValLeuThrValLeuGlu-AlaTrpAlaGln---------- 1580
                    |||   :::   |||   :::   :::   |||||||:::
Db   6812 GTCCATGGTGGACCCAGAGGCCTTGGCCATTGAGCAGAACCTGGGCAAAGCTCTAAAGAA 6871

Qy   1581 --------ProArgLysCysLeuProArgLeuGlnLeuLysAlaPro 1593
                  |||:::         |||:::      |||::::::      |||
Db   6872 AGATCTGAAAGAGAAGACCAAAACCAAAAAGCCAGGTACAAAGACCA 6918
```

Search completed: December 13, 2002, 00:45:32
Job time : 1027 secs

GenCore version 5.1.3

OM protein - nucleic search, using frame_plus_p2n model

```
Run on:         December 13, 2002, 00:24:46 ; Search time 124 Seconds
                                  (without alignments)
                                  3947.227 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters:      882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09902432/runat_12122002_141452_25831/app_query.fasta_1.1735
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09902432_@CGN_1_1_66_@runat_12122002_141452_25831 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Issued_Patents_NA:*
                1:  /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
                2:  /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
                3:  /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
                4:  /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
                5:  /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
                6:  /cgn2_6/ptodata/2/ina/backfiles1.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	6005	74.4	5134	2	US-08-635-121-1	Sequence 1, Appli
	2	4281	53.0	6605	1	US-08-769-309A-4	Sequence 4, Appli
	3	4281	53.0	6605	3	US-08-994-570-4	Sequence 4, Appli
	4	462	5.7	5361	4	US-08-973-462-2	Sequence 2, Appli
	5	462	5.7	6152	4	US-08-973-462-1	Sequence 1, Appli
	6	421.5	5.2	6744	1	US-08-119-125A-2	Sequence 2, Appli
	7	415	5.1	6306	1	US-08-466-390-3	Sequence 3, Appli
	8	415	5.1	6306	1	US-08-470-950-3	Sequence 3, Appli
	9	415	5.1	6306	1	US-08-467-781-3	Sequence 3, Appli
	10	415	5.1	6306	2	US-08-483-924-3	Sequence 3, Appli
	11	412	5.1	6306	1	US-08-195-487-3	Sequence 3, Appli
	12	412	5.1	6306	5	PCT-US93-06160-3	Sequence 3, Appli
	13	410.5	5.1	6755	3	US-08-931-999-4	Sequence 4, Appli
	14	388	4.8	11236	1	US-07-853-913-1	Sequence 1, Appli
	15	366	4.5	9551	1	US-08-056-200-93	Sequence 93, Appl
	16	366	4.5	9551	2	US-08-800-644-93	Sequence 93, Appl
	17	365.5	4.5	6414	4	US-09-134-001C-1626	Sequence 1626, Ap
	18	365.5	4.5	11091	4	US-09-134-001C-2243	Sequence 2243, Ap
	19	357.5	4.4	7766	4	US-09-125-619-3	Sequence 3, Appli
	20	354	4.4	8224	6	5180808-1	Patent No. 5180808
	21	351	4.3	5661	4	US-08-938-105-2	Sequence 2, Appli
c	22	346.5	4.3	32207	2	US-08-770-379-20	Sequence 20, Appl
c	23	346.5	4.3	32207	4	US-08-757-669A-20	Sequence 20, Appl
c	24	346.5	4.3	32207	4	US-09-230-371A-20	Sequence 20, Appl
	25	343.5	4.3	3337	1	US-08-072-610-1	Sequence 1, Appli
	26	343.5	4.3	3337	2	US-08-719-822B-1	Sequence 1, Appli
	27	343.5	4.3	3337	4	US-09-092-458-1	Sequence 1, Appli
c	28	339.5	4.2	3561	4	US-09-134-001C-1685	Sequence 1685, Ap
	29	334.5	4.1	3489	2	US-08-728-323A-1	Sequence 1, Appli
	30	334.5	4.1	3489	4	US-09-298-568-1	Sequence 1, Appli
	31	333.5	4.1	4852	1	US-07-853-913-3	Sequence 3, Appli
	32	321	4.0	5656	1	US-08-425-061-1	Sequence 1, Appli
	33	321	4.0	5656	1	US-08-825-886-1	Sequence 1, Appli
	34	321	4.0	5689	1	US-08-425-061-3	Sequence 3, Appli
	35	321	4.0	5689	1	US-08-825-886-3	Sequence 3, Appli
	36	321	4.0	5709	1	US-08-425-061-2	Sequence 2, Appli
	37	321	4.0	5709	1	US-08-825-886-2	Sequence 2, Appli
	38	321	4.0	5711	1	US-08-425-061-4	Sequence 4, Appli
	39	321	4.0	5711	1	US-08-825-886-4	Sequence 4, Appli
	40	321	4.0	5711	2	US-08-658-322-1	Sequence 1, Appli
	41	321	4.0	5712	1	US-08-425-061-12	Sequence 12, Appl
	42	321	4.0	5712	1	US-08-825-886-12	Sequence 12, Appl
	43	321	4.0	5712	2	US-08-603-753D-1	Sequence 1, Appli
	44	321	4.0	5712	3	US-09-099-753-1	Sequence 1, Appli
	45	321	4.0	5712	4	US-08-986-106-1	Sequence 1, Appli

ALIGNMENTS

```
RESULT 1
US-08-635-121-1
; Sequence 1, Application US/08635121
; Patent No. 5910442
;  GENERAL INFORMATION:
;    APPLICANT:  Gelman, Irwin H.
;    TITLE OF INVENTION:  TUMOR SUPPRESSOR GENE
;    NUMBER OF SEQUENCES:  2
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Brumbaugh, Graves, Donohue & Raymond
;      STREET:  30 Rockefeller Plaza
;      CITY:  New York
;      STATE:  NY
;      COUNTRY:  USA
;      ZIP:  10112-0228
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette
;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ Version 1.5
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/635,121
;      FILING DATE:  19-APRIL-1996
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Clark, Richard S
;      REGISTRATION NUMBER:  26,154
;      REFERENCE/DOCKET NUMBER:  A30558 - 165/33603
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  212-408-2558
;      TELEFAX:  212-765-2519
;      TELEX:
;  INFORMATION FOR SEQ ID NO:  1:
```

```
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  5134 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:
;    ORIGINAL SOURCE:
US-08-635-121-1

Alignment Scores:
Pred. No.:                0              Length:          5134
Score:                    6005.00        Matches:         1232
Percent Similarity:       97.33%         Conservative:    7
Best Local Similarity:    96.78%         Mismatches:      27
Query Match:              74.38%         Indels:          14
DB:                       2              Gaps:            4

US-09-902-432-4 (1-1596) x US-08-635-121-1 (1-5134)

Qy     329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
Db       2 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 61

Qy     349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
Db      62 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 121

Qy     369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
Db     122 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 181

Qy     389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
Db     182 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 241

Qy     409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
Db     242 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 301

Qy     429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
Db     302 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 361

Qy     449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
Db     362 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 421

Qy     469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
Db     422 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 481

Qy     489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
Db     482 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 541

Qy     509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
Db     542 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 601

Qy     529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
Db     602 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 661

Qy     549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568
Db     662 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 721

Qy     569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588
Db     722 GGGCCGCTGGAAGCACCCAG-GATGGGGAAGCTGAGGAAGGAACTACTTCGT--GGAGAG 778

Qy     589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608
Db     779 AAGAAGAGGAAGG--ATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 836

Qy     609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628
Db     837 CGGTC-CGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 895

Qy     629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648
Db     896 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 955

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668
Db     956 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 1015

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688
Db    1016 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATATAAGA 1075

Qy     689 GlyGlyProArgThrLeuGly-GlyAspSerHisArgAlaGluGluAlaSerLysAspLy 708
Db    1076 ---GGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAA 1132

Qy     708 sGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySe 728
Db    1133 AGAAGCCGA-ACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAG 1191

Qy     728 rSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSe 748
Db    1192 TTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTC 1251

Qy     748 rPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla-- 767
Db    1252 ATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA-AG 1310

Qy     768 ----GluAspSerSerVal---GluGlnLeuSerThrGluIleGluProSerArgGluGl 785
Db    1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy     785 uSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLy 805
Db    1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGAA 1430

Qy     805 s-GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProA 825
Db    1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy     825 snValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluA 845
Db    1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy     845 laGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865
Db    1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy     865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885
Db    1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy     885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905
Db    1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy     905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluG 925
Db    1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy     925 luThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValT 945
Db    1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847
```

```
Qy     945 hrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuA 965
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy     965 rgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaV 985
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy     985 alSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValG 1005
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy    1005 luSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaV 1025
           ||  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy    1025 alAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrG 1045
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy    1045 lySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerG 1065
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy    1065 luLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuA 1085
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy    1085 laGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrA 1105
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy    1105 laAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnA 1125
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy    1125 laValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProL 1145
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy    1145 euAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspS 1165
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy    1165 erLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaG 1185
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy    1185 lnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluG 1205
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy    1205 luProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProV 1225
           |||||||||||||||||||||||||||||||||||||||||    |||||||||||||||
Db    2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy    1225 alLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysV 1245
           ||    ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy    1245 alLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspV 1265
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy    1265 alThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluV 1285
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy    1285 alGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgG 1305
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy    1305 luThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisG 1325
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy    1325 luGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysA 1345
           ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy    1345 laLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleG 1365
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy    1365 luValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysV 1385
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy    1385 alIleGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisL 1405
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy    1405 euLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluA 1425
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy    1425 spThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProA 1445
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy    1445 laProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluA 1465
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy    1465 rgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrA 1485
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy    1485 laIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnL 1505
           |||||:::||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy    1505 ysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaP 1525
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy    1525 roGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgG 1545
           ||||||||||||||||||||||||||||||||||||||||     |||||||||||||||
Db    3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy    1545 luProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnP 1565
           ||||||||||||||||||||||||||||     |||||||||||||||||||||||||||
Db    3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy    1565 roArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysL 1585
           ||||||||||||||||||||||||||||||||||||||||||||          ||||||||||
Db    3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGC--TCAGCTCGGAAATGCT 3765

Qy    1585 euProArgLeuGlnLeuLysAlaProValSerLys 1596
           ||||||||||||||||||||||||||||||||||||
Db    3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 3800
```

RESULT 2
US-08-769-309A-4
; Sequence 4, Application US/08769309A
; Patent No. 5741890
; GENERAL INFORMATION:
;   APPLICANT: Scott, John D.,
;   APPLICANT: Nauert, Brian J.,
;   APPLICANT: Klauck, Theresa M.
;   TITLE OF INVENTION: Protein Binding Domains of Gravin
;   NUMBER OF SEQUENCES: 24
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
;   STREET: 6300 Sears Tower/233 South Wacker Drive
;   CITY: Chicago

```
;      STATE:  Illinois
;      COUNTRY:  United States of America
;      ZIP:  60606-6402
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/769,309A
;      FILING DATE:
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  No. 5741890and, Greta E.
;      REGISTRATION NUMBER:  35,302
;      REFERENCE/DOCKET NUMBER:  27866/33451
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-474-6300
;      TELEFAX:  312-474-0448
;      TELEX:  25-3856
;  INFORMATION FOR SEQ ID NO:  4:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6605 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  192..5531
US-08-769-309A-4

Alignment Scores:
Pred. No.:               1.51e-288      Length:        6605
Score:                   4281.00        Matches:       965
Percent Similarity:      68.69%         Conservative:  189
Best Local Similarity:   57.44%         Mismatches:    409
Query Match:             53.03%         Indels:        117
DB:                      1              Gaps:          37

US-09-902-432-4 (1-1596) x US-08-769-309A-4 (1-6605)

Qy       1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db     192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy      20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db     252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy      39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db     312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy      57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db     372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy      77 GlnGluGlyGln---------------------------------------GluGlu 82
Db     432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy      83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db     492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy     103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db     552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGCAGGAG 611

Qy     123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db     612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy     143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db     669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy     163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db     729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy     183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db     789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy     203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db     840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy     223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db     894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy     243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db     954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy     263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db    1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy     283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db    1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy     303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db    1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy     323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu-------------- 337
Db    1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy     338 -----------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db    1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy     354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db    1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy     374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db    1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy     393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db    1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy     413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db    1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy     433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db    1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy     453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db    1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy     473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db    1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706
```

```
Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTGCATCTGGTA---CCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3842

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
```

```
Db   3843 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3899

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
             |||   ||||||   |||    :::|||:::::          |||    :::|||
Db   3900 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3959

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
          :::||||||||      |||       |||   |||           ||||||   |||
Db   3960 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4013

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
                         |||      |||      :::   :::|||         :::
Db   4014 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4073

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
          :::|||      :::|||             |||         |||   |||::::::
Db   4074 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4133

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
            ::::::    |||:::||||||                |||      |||   |||
Db   4134 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4193

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
          |||      ||||||:::||||||   |||    :::|||       |||:::||||||
Db   4194 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4253

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
                |||   |||||||||               |||          :::   :::
Db   4254 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4313

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
          |||:::::::::|||       :::      |||   :::   ||||||   ||||||
Db   4314 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4373

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
          |||           |||::::::|||   |||   ::::::||||||||||||   :::
Db   4374 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4433

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
            :::   |||   ||||||   |||   |||:::   :::      ||||||
Db   4434 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4493

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
          |||   ||||||||||||:::   |||      ||||||||||||:::      ||||||
Db   4494 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4553

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
          ||||||            :::               ||||||   ||||||   |||
Db   4554 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4613

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
          |||:::   ||||||:::|||   |||:::||||||:::         :::      |||
Db   4614 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4673

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
                ||||||:::||||||    :::   |||   :::   :::|||:::|||   |||
Db   4674 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4733

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
          :::      |||      :::         |||   |||   |||||||||   :::
Db   4734 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4784

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
                |||||||||      |||||||||||||||:::||||||:::|||:::|||
Db   4785 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4838

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
          |||:::|||||||||||||||||||||||||   ||||||   |||||||||   |||
Db   4839 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4898

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
                :::   |||||||||               :::   ||||||::::::
Db   4899 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4958

Qy   1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
            :::      |||                      |||  :::|||||||||     ::
Db   4959 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5018

Qy   1562 lProGlnProArgGlu---------------------AspLeuGlnValLeuThrValLe 1575
            :   |||||||:::                     |||      :::|||
Db   5019 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5078

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
            :::          :::|||:::|||           |||:::   |||   |||:::
Db   5079 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5136
```

RESULT 3
US-08-994-570-4
; Sequence 4, Application US/08994570
; Patent No. 6090929
; GENERAL INFORMATION:
; APPLICANT: Scott, John D.,
; APPLICANT: Nauert, Brian J.,
; APPLICANT: Klauck, Theresa M.
; TITLE OF INVENTION: Protein Binding Domains of Gravin
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 6300 Sears Tower/233 South Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606-6402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/994,570
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090929and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 27866/33451
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6605 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 192..5531
US-08-994-570-4

```
Alignment Scores:
Pred. No.:              1.51e-288     Length:         6605
Score:                  4281.00       Matches:        965
Percent Similarity:     68.69%        Conservative:   189
Best Local Similarity:  57.44%        Mismatches:     409
Query Match:            53.03%        Indels:         117
DB:                     3             Gaps:           37
```

US-09-902-432-4 (1-1596) x US-08-994-570-4 (1-6605)

```
Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
          ||||||||||||||||||||||||||||||||||||||||||   |||   ||||||
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251
```

```
Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311
Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371
Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431
Qy     77 GlnGluGlyGln-----------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491
Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551
Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611
Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668
Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728
Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788
Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839
Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893
Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953
Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010
Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070
Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130
Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190
Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250
Qy    338 -----------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310
Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370
Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430
Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481
Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526
Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586
Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646
Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706
Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766
Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823
Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883
Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943
Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003
Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063
Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123
Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183
Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243
Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303
Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363
Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
```

```
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTGCATCTGGTA---CCAGTCAGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3842

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3843 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3899

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3900 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3959

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3960 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4013

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4014 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4073

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4074 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4133

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4134 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4193

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4194 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4253

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4254 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4313

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4314 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4373

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4374 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4433

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4434 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4493

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4494 ATTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4553
```

```
Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4554 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4613

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4614 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4673

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4674 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4733

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4734 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4784

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
Db   4785 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4838

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
Db   4839 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4898

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
Db   4899 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4958

Qy   1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
Db   4959 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5018

Qy   1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
Db   5019 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5078

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   5079 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5136
```

RESULT 4
US-08-973-462-2
; Sequence 2, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 5361
;   TYPE: DNA
;   ORGANISM: P. falciparum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(5361)
US-08-973-462-2

Alignment Scores:

Pred. No.:	8.51e-23	Length:	5361
Score:	462.00	Matches:	323
Percent Similarity:	36.36%	Conservative:	293
Best Local Similarity:	19.07%	Mismatches:	646
Query Match:	5.72%	Indels:	432
DB:	4	Gaps:	68

US-09-902-432-4 (1-1596) x US-08-973-462-2 (1-5361)

```
Qy     53 GlnLysAsnGlyGlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisVal 72
Db    283 CAAGTAAATGGTGAATTAGCTAGT---------------GAAGAAGTAAAGGAAAAAATT 327

Qy     73 GlnGluGluAsnGlnGluGlyGlnGlu---GluGluValValAspGluAspValGlyGln 91
Db    328 CTTGACTTATTAGAAGAAGGAAATACATTAACTGAAAGTGTAGATGATAATAAAAATTTA 387

Qy     92 ArgGluSerGluAspValArgGluLys---------AspArgValGluGlu--------- 105
Db    388 GAAGAAGCCGAAGATATAAAGGAAAATATCTTATTAAGTAATATAGAAGAACCAAAAGAA 447

Qy    106 --------------------------MetAlaAlaAsnSerThrAlaValGluAspIle 116
Db    448 AATATTATTGACAATTTATTAAATAATATTGGACAAAATTCAGAAAAACAAGAAAGTGTA 507

Qy    117 ThrLysAspGlyGln---------GluGluThrSerGluIleIleGluGlnIleProAla 133
Db    508 TCAGAAAATGTACAAGTCAGTGATGAACTTTTTAATGAATTATTAAATAGTGTAGATGTT 567

Qy    134 SerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGly 153
Db    568 AATGGAGAAGTAAAAGAAAATATTTTGGAG------GAAAGTCAAGTTAATGACGATATT 621

Qy    154 PheLysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGlu 173
Db    622 TTTAATAGTTTAGTAAAA------------------AGTGTTCAACAAGAACAACAACAC 663

Qy    174 LysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAla 193
Db    664 AATGTTGAAGAAAAAGTTGAAGAAAGTGTAGAAGAAAATGACGAAGAAAGTGTAGAAGAA 723

Qy    194 SerVal------GlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGlu 211
Db    724 AATGTAGAAGAAAATGTAGAAGAAAATGACGACGGAAGTGTAGCCTCAAGTGTTGAAGAA 783

Qy    212 Ser------AlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThr 229
Db    784 AGTATAGCTTCAAGTGTTGATGAAAGTATAGATTCAAGTATTGAAGAAAATGTAGCTCCA 843

Qy    230 LeuLysGlnGlu------------------------------------------------ 233
Db    844 ACTGTTGAAGAAATCGTAGCTCCAAGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAA 903

Qy    234 ------GlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGlu 251
Db    904 AGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAA 963

Qy    252 GluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSer 271
Db    964 AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAA 1023

Qy    272 ProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGly 291
Db   1024 ATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTGTAGCTCCA 1083

Qy    292 TrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGlu--- 310
Db   1084 AGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAA 1143

Qy    311 ---ThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGlu 329
Db   1144 AGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAA 1203

Qy    330 LysThrGluProAla---------SerGluGluGlnGluProAlaGluAspThrAspGln 346
Db   1204 AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAA 1263

Qy    347 AlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAsp 366
```

```
Db   1264 ATC---GTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATT------ 1314
Qy    367 LeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLys 386
Db   1315 GTAGCTCCAAGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAAT 1374
Qy    387 MetGluAlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGlu 406
Db   1375 GTTGAA------GAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAA 1428
Qy    407 GluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGlu 426
Db   1429 GAAAGT-------------------------------GTAGCTGAAAATGTTGAAGAA 1455
Qy    427 SerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAla 446
Db   1456 AGTGTA------------------------------------------GCTGAAAATGTT 1473
Qy    447 GluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThr 466
Db   1474 GAAGAAAGTGTAGCTGAA------------------------------------------ 1491
Qy    467 AspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluVal 486
Db   1492 ------AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTT 1545
Qy    487 GluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPhe 506
Db   1546 GAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTGTA 1605
Qy    507 SerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGly 526
Db   1606 GCTCCAAGTGTTGTAGAAAGTGTGGCT--------------------------------- 1632
Qy    527 GlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAsp 546
Db   1633 ------------CCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCT 1680
Qy    547 GluGlnLysGlyGluSerSerAlaSerSerProGluGluPro---------------Glu 561
Db   1681 GAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAA 1740
Qy    562 GluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGlu 581
Db   1741 GAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCT 1800
Qy    582 GlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLys 601
Db   1801 CCAACTGTT------------------GAAGAAATTGTAGCTCCA------AGTGTTGTA 1836
Qy    602 LysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGlu 621
Db   1837 GAAAGTGTGGCTCCAAGT------GTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGT 1890
Qy    622 Leu---GluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
Db   1891 GTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAAT 1950
Qy    641 GlnAspGluVal------LysThrValGlyGluGluGlnLysProGluGluProLysArg 658
Db   1951 GTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCA--------------- 1995
Qy    659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
Db   1996 ------------------ACTGTTGAAGAAATTGTAGCTCCAAGTGTTGTAGAAAGTGTG 2037
Qy    679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
Db   2038 GCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTA------------ 2085
Qy    699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
Db   2086 ---------------GCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGT 2130
Qy    719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
Db   2131 GTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACT 2190
Qy    739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPro--------- 755
Db   2191 GTTGAAGAAATCGTAGCT------CCAACTGTTGAAGAAATTGTAGCTCCAAGTGTTGTA 2244
Qy    756 ---------ArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerVal 772
Db   2245 GAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCT 2304
Qy    773 GluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLys 792
Db   2305 GAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGT------------------ 2346
Qy    793 PheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAla------Thr 810
Db   2347 ------------------------GTAGCTGAAAATGTTGAAGAAAGTGTAGCTCCAACT 2382
Qy    811 ValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValVal 830
Db   2383 GTTGAAGAAATTGTAGCTCCAAGTGTTGAAGAA------------------AGTGTAGCT 2424
Qy    831 ProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGlu 850
Db   2425 CCAAGTGTTGAAGAAAGTGTTGCTGAAAACGTTGCAACAAATTTATCAGACAATCTT--- 2481
Qy    851 LeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHis 870
Db   2482 TTAAGTAATTTATTAGGTGGTATC---GAAACTGAGGAAATAAAGGACAGTATATTAAAT 2538
Qy    871 ThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSer 890
Db   2539 GAGATAGAA--------------------------------GAAGTAAAAGAAAAT--- 2562
Qy    891 ProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMet 910
Db   2563 ---------GTAGTCACCACAATACTAGAAAACGTAGAAGAAACTACAGCTGAAAGTGTA 2613
Qy    911 ProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProValLeuThr 930
Db   2614 ACTACTTTTAGTAACATATTAGAGGAG---------ATACAAGAAAATACTATTACTAAT 2664
Qy    931 GlnThrLeuProGluGly---LysAspAlaHisAspAspMetValThrSerGluValAsp 949
Db   2665 GATACTATAGAGGAAAAATTAGAAGAACTCCACGAAAATGTATTAAGTGCCGCTTTAGAA 2724
Qy    950 PheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluVal 969
Db   2725 AATACCCAAAGTGAAGAGGAAAAGAAAGAAGTAATAGATGTAATT------GAAGAAGTA 2778
Qy    970 ThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGln------ 987
Db   2779 AAAGAA------------GAGGTCGCTACCACTTTAATAGAAACTGTGGAACAGGCAGAA 2826
Qy    988 ------------------LeuThrAspSerProAspThrThrGluGluAlaThrProVal 1001
Db   2827 GAAAAGAGCGCAAATACAATTACGGAAATATTTGAAAATTTAGAAGAA------------ 2874
Qy   1002 GlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIle 1021
Db   2875 AATGCAGTAGAAAGTAATGAAAATGTTGCAGAGAATTTAGAGAAATTAAACGAAACTGTA 2934
Qy   1022 LeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrVal 1041
Db   2935 TTTAATACTGTATTAGATAAAGTAGAGGAAACAGTAGAAATTAGCGGAGAAAGTTTAGAA 2994
Qy   1042 GlnArgThrGlySerLysAla---------------LeuGluLysValGluGluValGlu 1056
Db   2995 AACAATGAAATGGATAAAGCATTTTTAGTGAAATATTTGATAATGTAAAAGGAATACAA 3054
```

```
Qy   1057 GluAsp--------------------------------SerGluValLeuAlaSerGlu 1065
Db   3055 GAAAATTTATTAACAGGTATGTTTCGAAGTATAGAAACCAGTATAGTAATCCAATCAGAA 3114

Qy   1066 LysGluLysAspVal-----------------------------------------Met 1071
Db   3115 GAAAAGGTTGATTTGAATGAAAATGTGGTTAGTTCGATTTTAGATAATATAGAAAATATG 3174

Qy   1072 ProLysGly----------------------------------ProValGlnGluAla 1079
Db   3175 AAAGAAGGTTTATTAAATAAATTAGAAAATATTTCAAGTACTGAAGGTGTTCAAGAAACT 3234

Qy   1080 GlyAlaGluHisLeuAlaGlnGly---------------------------------- 1087
Db   3235 GTAACTGAACATGTAGAACAAAATGTATATGTGGATGTTGATGTTCCTGCTATGAAAGAT 3294

Qy   1088 -----------------SerGluThrGlyGlnAlaThrProGluSerLeuGluValPro 1101
Db   3295 CAATTTTTAGGAATATTAAATGAGGCAGGAGGGTTGAAAGAAATGTTTTTAATTTGGAA 3354

Qy   1102 GluValThrAlaAspValAspHisValAlaThrCysGlnValIleLys---------Leu 1118
Db   3355 GATGTATTTAAAAGTGAAAGTGATGTAATTACTGTAGAAGAAATTAAGGATGAACCGGTT 3414

Qy   1119 GlnGlnLeuMetGluGln---------AlaValAlaProGluSerSerGluThrLeuThr 1135
Db   3415 CAAAAAGAGGTAGAAAAAGAAACTGTTAGTATTATTGAAGAAATGGAAGAAAATATTGTA 3474

Qy   1136 AspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGln 1155
Db   3475 GATGTATTAGAGGAAGAAAAAGAAGATTTAACAGACAAG---ATGATAGATGCAGTAGAA 3531

Qy   1156 GlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGln 1175
Db   3532 GAATCCATAGAAATATCTTCAGATTCTAAAGAAGAAACTGAATCTATTAAAGATAAA--- 3588

Qy   1176 ValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsn 1195
Db   3589 ------GAAAAAGATGTTTCACTAGTTGTTGAAGAA----------------------- 3618

Qy   1196 ValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGlu------Pro 1213
Db   3619 ------GTTCAAGACAATGATATGGATGAAAGTGTTGAGAAAGTTTTAGAATTGAAAAAT 3672

Qy   1214 ThrGlnGlnGluLeuThrAlaAlaAlaValProVal--------------------Leu 1226
Db   3673 ATGGAAGAGGAGTTAATGAAGGATGCTGTTGAAATAAATGACATTACTAGCAAACTTATT 3732

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu---AspGlyGluLysVal 1245
Db   3733 GAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATAAAAGATATGGAAAAATTA 3792

Qy   1246 LysGlu-----------------------GluGlnGluValPheValHisSerGlyPro 1257
Db   3793 AAAGAATTAGAAAAAGCATTATCAGAAGATTCTAAAGAAATAATAGATGCAAAAGATGAT 3852

Qy   1258 AsnSerGlnLysAlaAlaAspValThrTyrAsp---------------SerGluValMet 1272
Db   3853 ACATTAGAAAAAGTTATTGAAGAGGAACATGATATAACGACGACGTTGGATGAAGTTGTA 3912

Qy   1273 GlyValAlaGlyCysGlnGlu------------------------------------- 1279
Db   3913 GAATTAAAAGATGTCGAAGAAGACAAGATCGAAAAAGTATCTGATTTAAAAGATCTTGAA 3972

Qy   1280 --------------------LysGluSerThrGluValGlnSerLeuSerLeuGluGlu 1292
Db   3973 GAAGATATATTAAAAGAAGTAAAAGAAATCAAAGAACTTGAAAGTGAAATTTTAGAAGAT 4032

Qy   1293 GlyGlu-----------MetGluThrAspValGluLysGluLysArgGluThrLysPro 1308
Db   4033 TATAAAGAATTAAAAACTATTGAAACAGATATTTTAGAAGAGAAAAAAGAAATAGAAAAA 4092

Qy   1309 GluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyr 1328
Db   4093 GATCATTTTGAAAAATTCGAAGAAGAAGCTGAAGAAATAAAAGATCTTGAAGCAGATATA 4152

Qy   1329 GlyLysProValLeuThrLeuAspMetProSerSerGlu-------------------- 1341
Db   4153 TTAAAAGAAGTATCTTCATTAGAAGTTGAAGAAGAAAAAAAATTAGAAGAAGTACACGAA 4212

Qy   1341 ----------------------------------------------------------- 1341
Db   4213 TTAAAAGAAGAGGTAGAACATATAATAAGTGGTGATGCGCATATAAAAGGTTTGGAAGAA 4272

Qy   1342 --------------------------ArgGlyLysAlaLeuGlySerLeuGlyGlySer 1352
Db   4273 GATGATTTAGAAGAAGTAGATGATTTAAAAGGAAGTATATTAGACATGTTAAAGGGAGAT 4332

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4333 ATGGAATTAGGGGATATGGATAAG-----------------------GAAAGTTTAGAA 4368

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrValValIleSer 1392
Db   4369 GATGTAACAACAAAACTTGGAGAAAGAGTTGAATCC---TTAAAAGATGTTTTATCTAGT 4425

Qy   1393 GluThrGlyGluSerProGluCysValGlyAlaHis----------------------- 1404
Db   4426 GCATTAGGCATGGATGAAGAACAAATGAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTG 4485

Qy   1405 ---------LeuLeu--------------------ProAlaGluLysSerSerAlaThr 1414
Db   4486 GAAGAAGTATTATTAAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAATAACAAAAAAG 4545

Qy   1415 GlyGlyHisTrpThrLeuGlnHisAlaGlu------AspThrValProLeuGlyProGlu 1432
Db   4546 AAAGTAAGGTTTGATATTAAGGATAAGGAACCAAAAGATGAAATA---GTAGAAGTTGAA 4602

Qy   1433 SerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisPro 1452
Db   4603 ATGAAAGATGAAGATATA-----------------------GAAGAAGATGTAGAAGAA 4638

Qy   1453 AspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysPro 1472
Db   4639 GATATAGAAGAAGATATAGAAGAAGATAAAGTTGAAGATATAGATGAAGATATAGATGAA 4698

Qy   1473 AspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeuLysAla 1492
Db   4699 GATATAGGTGAAGACAAAGAT------------GAAGTTATAGATTTAATAGTCCAAAAA 4746

Qy   1493 GluProGluIleLeuGluLeuGluSerLysSerAsnLysIle 1506
Db   4747 GAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAAAATTA 4788
```

RESULT 5
US-08-973-462-1
; Sequence 1, Application US/08973462B
; Patent No. 6191270
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 0660-0125-0 PCT
; CURRENT APPLICATION NUMBER: US/08/973,462B
; CURRENT FILING DATE: 1998-02-06
; EARLIER APPLICATION NUMBER: PCT/FR96/00894
; EARLIER FILING DATE: 1996-06-12
; EARLIER APPLICATION NUMBER: FR 95/07007
; EARLIER FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 6152
; TYPE: DNA
; ORGANISM: P. falciparum
US-08-973-462-1

```
Alignment Scores:
Pred. No.:              1.02e-22      Length:         6152
Score:                  462.00        Matches:        323
Percent Similarity:     36.36%        Conservative:   293
Best Local Similarity:  19.07%        Mismatches:     646
Query Match:            5.72%         Indels:         432
DB:                     4             Gaps:           68

US-09-902-432-4 (1-1596) x US-08-973-462-1 (1-6152)

Qy     53 GlnLysAsnGlyGlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisVal 72
Db    527 CAAGTAAATGGTGAATTAGCTAGT---------------GAAGAAGTAAAGGAAAAAATT 571

Qy     73 GlnGluGluAsnGlnGluGlyGlnGlu---GluGluValValAspGluAspValGlyGln 91
Db    572 CTTGACTTATTAGAAGAAGGAAATACATTAACTGAAAGTGTAGATGATAATAAAAATTTA 631

Qy     92 ArgGluSerGluAspValArgGluLys---------AspArgValGluGlu--------- 105
Db    632 GAAGAAGCCGAAGATATAAAGGAAAATATCTTATTAAGTAATATAGAAGAACCAAAAGAA 691

Qy    106 ---------------------------MetAlaAlaAsnSerThrAlaValGluAspIle 116
Db    692 AATATTATTGACAATTTATTAAATAATATTGGACAAAATTCAGAAAAACAAGAAAGTGTA 751

Qy    117 ThrLysAspGlyGln---------GluGluThrSerGluIleIleGluGlnIleProAla 133
Db    752 TCAGAAAATGTACAAGTCAGTGATGAACTTTTTAATGAATTATTAAATAGTGTAGATGTT 811

Qy    134 SerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGly 153
Db    812 AATGGAGAAGTAAAAGAAAATATTTTGGAG------GAAAGTCAAGTTAATGACGATATT 865

Qy    154 PheLysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGlu 173
Db    866 TTTAATAGTTTAGTAAAA-----------------AGTGTTCAACAAGAACAACAACAC 907

Qy    174 LysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAla 193
Db    908 AATGTTGAAGAAAAAGTTGAAGAAAGTGTAGAAGAAAATGACGAAGAAAGTGTAGAAGAA 967

Qy    194 SerVal------GlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGlu 211
Db    968 AATGTAGAAGAAAATGTAGAAGAAAATGACGACGGAAGTGTAGCCTCAAGTGTTGAAGAA 1027

Qy    212 Ser------AlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThr 229
Db   1028 AGTATAGCTTCAAGTGTTGATGAAAGTATAGATTCAAGTATTGAAGAAAATGTAGCTCCA 1087

Qy    230 LeuLysGlnGlu------------------------------------------------ 233
Db   1088 ACTGTTGAAGAAATCGTAGCTCCAAGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAA 1147

Qy    234 ------GlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGlu 251
Db   1148 AGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAA 1207

Qy    252 GluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSer 271
Db   1208 AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAA 1267

Qy    272 ProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGly 291
Db   1268 ATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTGTAGCTCCA 1327

Qy    292 TrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGlu--- 310
Db   1328 AGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAA 1387

Qy    311 ---ThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGlu 329
Db   1388 AGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAA 1447

Qy    330 LysThrGluProAla---------SerGluGluGlnGluProAlaGluAspThrAspGln 346
Db   1448 AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAA 1507

Qy    347 AlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAsp 366
Db   1508 ATC---GTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATT------ 1558

Qy    367 LeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLys 386
Db   1559 GTAGCTCCAAGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAAT 1618

Qy    387 MetGluAlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGlu 406
Db   1619 GTTGAA------GAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAA 1672

Qy    407 GluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGlu 426
Db   1673 GAAAGT------------------------------GTAGCTGAAAATGTTGAAGAA 1699

Qy    427 SerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAla 446
Db   1700 AGTGTA------------------------------------------GCTGAAAATGTT 1717

Qy    447 GluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThr 466
Db   1718 GAAGAAAGTGTAGCTGAA------------------------------------------ 1735

Qy    467 AspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluVal 486
Db   1736 ------AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTT 1789

Qy    487 GluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPhe 506
Db   1790 GAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTGTA 1849

Qy    507 SerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGly 526
Db   1850 GCTCCAAGTGTTGTAGAAAGTGTGGCT--------------------------------- 1876

Qy    527 GlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAsp 546
Db   1877 ------------CCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCT 1924

Qy    547 GluGlnLysGlyGluSerSerAlaSerSerProGluGluPro---------------Glu 561
Db   1925 GAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAA 1984

Qy    562 GluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGlu 581
Db   1985 GAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCT 2044

Qy    582 GlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLys 601
Db   2045 CCAACTGTT------------------GAAGAAATTGTAGCTCCA------AGTGTTGTA 2080

Qy    602 LysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGlu 621
Db   2081 GAAAGTGTGGCTCCAAGT------GTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGT 2134

Qy    622 Leu---GluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
Db   2135 GTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAAT 2194

Qy    641 GlnAspGluVal------LysThrValGlyGluGluGlnLysProGluGluProLysArg 658
Db   2195 GTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCA--------------- 2239

Qy    659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
Db   2240 ------------------ACTGTTGAAGAAATTGTAGCTCCAAGTGTTGTAGAAAGTGTG 2281
```

```
Qy    679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
Db   2282 GCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTA------------ 2329

Qy    699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
Db   2330 ---------------GCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGT 2374

Qy    719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
Db   2375 GTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACT 2434

Qy    739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPro--------- 755
Db   2435 GTTGAAGAAATCGTAGCT------CCAACTGTTGAAGAAATTGTAGCTCCAAGTGTTGTA 2488

Qy    756 ---------ArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerVal 772
Db   2489 GAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCT 2548

Qy    773 GluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLys 792
Db   2549 GAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGT------------------ 2590

Qy    793 PheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAla------Thr 810
Db   2591 -----------------------GTAGCTGAAAATGTTGAAGAAAGTGTAGCTCCAACT 2626

Qy    811 ValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValVal 830
Db   2627 GTTGAAGAAATTGTAGCTCCAAGTGTTGAAGAA-----------------AGTGTAGCT 2668

Qy    831 ProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGlu 850
Db   2669 CCAAGTGTTGAAGAAAGTGTTGCTGAAAACGTTGCAACAAATTTATCAGACAATCTT--- 2725

Qy    851 LeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHis 870
Db   2726 TTAAGTAATTTATTAGGTGGTATC---GAAACTGAGGAAATAAAGGACAGTATATTAAAT 2782

Qy    871 ThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSer 890
Db   2783 GAGATAGAA--------------------------------GAAGTAAAAGAAAAT--- 2806

Qy    891 ProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMet 910
Db   2807 ---------GTAGTCACCACAATACTAGAAAACGTAGAAGAAACTACAGCTGAAAGTGTA 2857

Qy    911 ProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProValLeuThr 930
Db   2858 ACTACTTTTAGTAACATATTAGAGGAG---------ATACAAGAAAATACTATTACTAAT 2908

Qy    931 GlnThrLeuProGluGly---LysAspAlaHisAspAspMetValThrSerGluValAsp 949
Db   2909 GATACTATAGAGGAAAAATTAGAAGAACTCCACGAAAATGTATTAAGTGCCGCTTTAGAA 2968

Qy    950 PheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluVal 969
Db   2969 AATACCCAAAGTGAAGAGGAAAAGAAAGAAGTAATAGATGTAATT------GAAGAAGTA 3022

Qy    970 ThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGln------ 987
Db   3023 AAAGAA------------GAGGTCGCTACCACTTTAATAGAAACTGTGGAACAGGCAGAA 3070

Qy    988 ------------------LeuThrAspSerProAspThrThrGluGluAlaThrProVal 1001
Db   3071 GAAAAGAGCGCAAATACAATTACGGAAATATTTGAAAATTTAGAAGAA------------ 3118

Qy   1002 GlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIle 1021
Db   3119 AATGCAGTAGAAAGTAATGAAAATGTTGCAGAGAATTTAGAGAAATTAAACGAAACTGTA 3178

Qy   1022 LeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrVal 1041
Db   3179 TTTAATACTGTATTAGATAAAGTAGAGGAAACAGTAGAAATTAGCGGAGAAAGTTTAGAA 3238

Qy   1042 GlnArgThrGlySerLysAla---------------LeuGluLysValGluGluValGlu 1056
Db   3239 AACAATGAAATGGATAAAGCATTTTTTAGTGAAATATTTGATAATGTAAAAGGAATACAA 3298

Qy   1057 GluAsp------------------------------SerGluValLeuAlaSerGlu 1065
Db   3299 GAAAATTTATTAACAGGTATGTTTCGAAGTATAGAAACCAGTATAGTAATCCAATCAGAA 3358

Qy   1066 LysGluLysAspVal------------------------------------------Met 1071
Db   3359 GAAAAGGTTGATTTGAATGAAAATGTGGTTAGTTCGATTTTAGATAATATAGAAAATATG 3418

Qy   1072 ProLysGly------------------------------------ProValGlnGluAla 1079
Db   3419 AAAGAAGGTTTATTAAATAAATTAGAAAATATTTCAAGTACTGAAGGTGTTCAAGAAACT 3478

Qy   1080 GlyAlaGluHisLeuAlaGlnGly------------------------------------ 1087
Db   3479 GTAACTGAACATGTAGAACAAAATGTATATGTGGATGTTGATGTTCCTGCTATGAAAGAT 3538

Qy   1088 ------------------SerGluThrGlyGlnAlaThrProGluSerLeuGluValPro 1101
Db   3539 CAATTTTTAGGAATATTAAATGAGGCAGGAGGGTTGAAAGAAATGTTTTTAATTTGGAA 3598

Qy   1102 GluValThrAlaAspValAspHisValAlaThrCysGlnValIleLys---------Leu 1118
Db   3599 GATGTATTTAAAAGTGAAAGTGATGTAATTACTGTAGAAGAAATTAAGGATGAACCGGTT 3658

Qy   1119 GlnGlnLeuMetGluGln---------AlaValAlaProGluSerSerGluThrLeuThr 1135
Db   3659 CAAAAAGAGGTAGAAAAAGAAACTGTTAGTATTATTGAAGAAATGGAAGAAAATATTGTA 3718

Qy   1136 AspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGln 1155
Db   3719 GATGTATTAGAGGAAGAAAAAGAAGATTTAACAGACAAG---ATGATAGATGCAGTAGAA 3775

Qy   1156 GlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGln 1175
Db   3776 GAATCCATAGAAATATCTTCAGATTCTAAAGAAGAAACTGAATCTATTAAAGATAAA--- 3832

Qy   1176 ValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsn 1195
Db   3833 ------GAAAAAGATGTTTCACTAGTTGTTGAAGAA------------------------ 3862

Qy   1196 ValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGlu------Pro 1213
Db   3863 ------GTTCAAGACAATGATATGGATGAAAGTGTTGAGAAAGTTTTAGAATTGAAAAAT 3916

Qy   1214 ThrGlnGlnGluLeuThrAlaAlaAlaValProVal---------------------Leu 1226
Db   3917 ATGGAAGAGGAGTTAATGAAGGATGCTGTTGAAATAAATGACATTACTAGCAAACTTATT 3976

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu---AspGlyGluLysVal 1245
Db   3977 GAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATAAAAGATATGGAAAAATTA 4036

Qy   1246 LysGlu-----------------------GluGlnGluValPheValHisSerGlyPro 1257
Db   4037 AAAGAATTAGAAAAAGCATTATCAGAAGATTCTAAAGAAATAATAGATGCAAAAGATGAT 4096

Qy   1258 AsnSerGlnLysAlaAlaAspValThrTyrAsp---------------SerGluValMet 1272
Db   4097 ACATTAGAAAAAGTTATTGAAGAGGAACATGATATAACGACGACGTTGGATGAAGTTGTA 4156

Qy   1273 GlyValAlaGlyCysGlnGlu--------------------------------------- 1279
Db   4157 GAATTAAAAGATGTCGAAGAAGACAAGATCGAAAAAGTATCTGATTTAAAAGATCTTGAA 4216

Qy   1280 ---------------------LysGluSerThrGluValGlnSerLeuSerLeuGluGlu 1292
```

```
                                 ||||||        |||:::::|||       ||||||:::
Db   4217 GAAGATATATTAAAAGAAGTAAAAGAAATCAAAGAACTTGAAAGTGAAATTTTAGAAGAT 4276

Qy   1293 GlyGlu-----------MetGluThrAspValGluLysGluLysArgGluThrLysPro 1308
             :::              :::|||||||||:::   :::||||||:::|||    :::
Db   4277 TATAAAGAATTAAAAACTATTGAAACAGATATTTTAGAAGAGAAAAAAGAAATAGAAAAA 4336

Qy   1309 GluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyr 1328
             :::        :::      |||:::                :::   |||
Db   4337 GATCATTTTGAAAAATTCGAAGAAGAAGCTGAAGAAATAAAAGATCTTGAAGCAGATATA 4396

Qy   1329 GlyLysProValLeuThrLeuAspMetProSerSerGlu-------------------- 1341
                |||    |||   :::|||:::::          :::
Db   4397 TTAAAAGAAGTATCTTCATTAGAAGTTGAAGAAGAAAAAAAATTAGAAGAAGTACACGAA 4456

Qy   1341 ------------------------------------------------------------ 1341

Db   4457 TTAAAAGAAGAGGTAGAACATATAATAAGTGGTGATGCGCATATAAAAGGTTTGGAAGAA 4516

Qy   1342 --------------------------ArgGlyLysAlaLeuGlySerLeuGlyGlySer 1352
                                    :::|||      |||      |||    |||
Db   4517 GATGATTTAGAAGAAGTAGATGATTTAAAAGGAAGTATATTAGACATGTTAAAGGGAGAT 4576

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
                |||    |||    ||||||                        :::||||||:::
Db   4577 ATGGAATTAGGGGATATGGATAAG-----------------------GAAAGTTTAGAA 4612

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrValValIleSer 1392
                   |||:::       |||   ||||||        ::::::   |||:::    |||
Db   4613 GATGTAACAACAAAACTTGGAGAAAGAGTTGAATCC---TTAAAAGATGTTTTATCTAGT 4669

Qy   1393 GluThrGlyGluSerProGluCysValGlyAlaHis----------------------- 1404
                |||          |||    :::
Db   4670 GCATTAGGCATGGATGAAGAACAAATGAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTG 4729

Qy   1405 ---------LeuLeu--------------------ProAlaGluLysSerSerAlaThr 1414
                   ||||||                     |||    :::|||    :::
Db   4730 GAAGAAGTATTATTAAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAATAACAAAAAAG 4789

Qy   1415 GlyGlyHisTrpThrLeuGlnHisAlaGlu------AspThrValProLeuGlyProGlu 1432
                   :::    ::::::         |||       |||    :::    :::       |||
Db   4790 AAAGTAAGGTTTGATATTAAGGATAAGGAACCAAAAGATGAAATA---GTAGAAGTTGAA 4846

Qy   1433 SerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisPro 1452
             :::    |||    |||                               |||       :::
Db   4847 ATGAAAGATGAAGATATA------------------------GAAGAAGATGTAGAAGAA 4882

Qy   1453 AspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysPro 1472
          |||::::::    :::|||          :::     |||         :::|||:::
Db   4883 GATATAGAAGAAGATATAGAAGAAGATAAAGTTGAAGATATAGATGAAGATATAGATGAA 4942

Qy   1473 AspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeuLysAla 1492
          |||    |||    |||    |||                  |||:::    :::::::::
Db   4943 GATATAGGTGAAGACAAAGAT------------GAAGTTATAGATTTAATAGTCCAAAAA 4990

Qy   1493 GluProGluIleLeuGluLeuGluSerLysSerAsnLysIle 1506
          |||       |||    ::::::::::::|||        |||:::
Db   4991 GAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAAAATTA 5032
```

RESULT 6
US-08-119-125A-2

```
; Sequence 2, Application US/08119125A
; Patent No. 5610011
;  GENERAL INFORMATION:
;    APPLICANT:  SMITH, Hilda Elizabeth
;    APPLICANT:  VECHT, Uri
;    TITLE OF INVENTION:  DNA Sequences which code for Virulence
;    TITLE OF INVENTION:  Characteristics of Streptococcus suis and parts thereof, polyp
;    TITLE OF INVENTION:  antibodies derived therefrom and the use thereof for the diagn
;    TITLE OF INVENTION:  protection against infection by S. suis in mammals, including
;    NUMBER OF SEQUENCES:  3
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Centraal Diergeneeskundig Instituut
;      STREET:  Edelhertweg 15
;      CITY:  PH Lelystad
;      STATE:
;      COUNTRY:  The Netherlands
;      ZIP:  NL-8219
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.44 MB storage
;      COMPUTER:  IBM  compatible
;      OPERATING SYSTEM:  MS-DOS  v.6.0
;      SOFTWARE:  WordPerfect  v. 6.0
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/119,125A
;      FILING DATE:  20-SEP-1993
;      CLASSIFICATION:  514
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/NL92/00054
;      FILING DATE:  19-MAR-1992
;      APPLICATION NUMBER:  NL 9100510
;      FILING DATE:  21-MAR-1991
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Handal, Anthony H.
;      REGISTRATION NUMBER:  26275
;      REFERENCE/DOCKET NUMBER:  SMITHHE119125
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (203) 838-8589
;      TELEFAX:  (203) 838-8794
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6744 base pairs
;      TYPE:  Nucleic acid with corresponding amino acids
;      STRANDEDNESS:  single stranded
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  genomic DNA
;    ORIGINAL SOURCE:
;      ORGANISM:  Streptococcus suis type II (pathogenic)
;    FEATURE:
;      OTHER INFORMATION:  Extracellular factor related protein (EF*) gene
;    FEATURE:
;      NAME/KEY:  promoter -35 region
;      LOCATION:  bp 66 to 71
;    FEATURE:
;      NAME/KEY:  promoter -10 region
;      LOCATION:  bp 89 to 94
;    FEATURE:
;      NAME/KEY:  promoter -35 region
;      LOCATION:  bp 153 to 158
;    FEATURE:
;      NAME/KEY:  promoter -10 region
;      LOCATION:  bp 176 to 181
;    FEATURE:
;      NAME/KEY:  ribosome binding site
;      LOCATION:  bp 350 to 356
;    FEATURE:
;      NAME/KEY:  signal peptide
;      LOCATION:  bp 361 to 498
;    FEATURE:
;      NAME/KEY:  start of repetitive units R1-R11
;      LOCATION:  bp 2869, 3097, 3292, 3520, 4087, 4381, 4609, 4837,
;      LOCATION:  5065, 5293, 5521:
;    FEATURE:
;      NAME/KEY:  start of repetitive Asn-Pro-Asn-Leu sequences
;      LOCATION:  bp 2932, 3160, 3355, 3583, 4150, 4444, 4672, 4900,
;      LOCATION:  5128, 5356, 5584:
;    FEATURE:
;      NAME/KEY:  dyad symmetry regions
;      LOCATION:  from bp 6554 to 6566 and from bp 6571 to 6583
;    FEATURE:
;      NAME/KEY:  dyad symmetry regions
;      LOCATION:  from bp 6611 to 6625 and from bp 6631 to 6644
US-08-119-125A-2

Alignment Scores:
Pred. No.:            7.62e-20       Length:        6744
```

```
Score:                    421.50     Matches:        333
Percent Similarity:       36.25%     Conservative:   268
Best Local Similarity:    20.08%     Mismatches:     664
Query Match:              5.22%      Indels:         393
DB:                       1          Gaps:           71

US-09-902-432-4 (1-1596) x US-08-119-125A-2 (1-6744)

Qy      4 GlySerSerThrGluGlnArgSerProGluGlnProAlaGly------------------ 17
          |||::::::|||:::       |||        |||  |||
Db   2056 GGTACAAACACTAAGATTACTTCAGAACTGGATCCTACGGGAGCAACTGCAAACCAAGGA 2115

Qy     18 ------------SerAspThrProSerGluLeuValLeuSerGlyHisGlyProAlaAla 33
                      |||   |||           :::   :::::|||   ||||||||
Db   2116 GATGACGGTCAATCTTCAACTAAGTTTAACGTTAAGATTACAGGTACCGGACCTGCTACA 2175

Qy     34 GluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspProAlaThrLysLeuProGln 53
          |||   :::|||
Db   2176 GAAGGTACCGGCACTTAT------------------------------------------ 2193

Qy     54 LysAsnGlyGlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGln 73
                                                         :::   |||
Db   2194 ---------------------------------------------AAGCTTCGTGTTGGA 2208

Qy     74 GluGluAsnGlnGlu---GlyGlnGluGluGluValValAsp------------GluAsp 88
          |||:::|||         |||   |||   ::::::|||||             |||:::
Db   2209 GAAGATAACTATCCTTTTGGTCCAGAGGGGAAACTTGTTGATGGAAATAAACCAGAAAAT 2268

Qy     89 ValGlyGlnArgGluSerGluAspValArgGluLysAspArgValGluGluMetAlaAla 108
          ||||||                  :::            |||                :::
Db   2269 GTAGGTTTGACATCTGTAAAAGTTACCTTCGTAAAA---------CATGCTACGGTGTCA 2319

Qy    109 AsnSerThrAlaValGluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIle 128
                :::||||||:::                     ||||||   :::   :::|||
Db   2320 ACACCAGTTTCTGTTGAAAATCCAGCTAACTTAACGCCAGAAGAAAAAGCCGCAGTTATT 2379

Qy    129 GluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnProAlaGluSerGln 148
             ||||||               :::   |||   :::            :::|||
Db   2380 GCTCAAATCAAGAAAGACAACGCAGACAACGAAAGATTGAAGGGCTTGCCAGATTCA--- 2436

Qy    149 AlaAsnAspValGlyPheLysLysValPheLysPheValGlyPheLysPheThrValLys 168
                                                             |||||||||
Db   2437 ---------------------------------------------GCATTTACAGTTAAC 2451

Qy    169 LysAspLysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGly 188
             |||                                     ||||||      |||
Db   2452 TCAGATGGT------------------------------ACTGTGTCAGTTGACTACAGT 2481

Qy    189 GluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAla 208
             |||               ||||||   |||         :::   :::
Db   2482 GCCGGTGGTGTCAATGTTGATGGTGCGACAGACATTATTAAGAATGCTACCACAAACTTG 2541

Qy    209 ValGly------GluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGln 226
                         |||      |||   :::   |||   :::      |||:::
Db   2542 GCAGATACACGGAATGAAGCAAAAGCAGAAATCGACACAAAATTAGCTGAACATAAAAAA 2601

Qy    227 GluGlyThrLeuLysGlnGluGlnSer------------SerThrGluIleProLeuGln 242
                   |||::::::::::::                     :::|||   |||:::
Db   2602 GCTATCGAAGCAAAACGGGATGAAGCGTTTTCTAAAATTGATGATGACATTTCCTTGAGA 2661

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
          ||||||         |||||||||::::::                     :::   :::
Db   2662 GCAGAACAGAGACAGGCTGCTAAGGATGCCGTTGCTGCAGCTGCTGGGGATGCTTTGAAA 2721

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
          :::            |||:::   |||:::               :::   |||   |||
Db   2722 GAATTAGACACAACAAGGCGACAGAAGCAAAAGAAAAAATTGATAAAGCTACGACGGCCTCA 2781

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLys--------------- 297
                            |||:::|||
Db   2782 GAAATCAATGATGCTAAGACTAATGGTGAGATTAATCTGGACAGTGCAGAAGCAGTAGGC 2841

Qy    298 LysThrSerPheLysLysSerLysGluAspAspLeuGluThrAlaGlu---LysArgLys 316
          :::   :::         ::::::||||||   :::|||         ||||||   :::   |||
Db   2842 GAAAAAGCTATTAACCAGGCGAAGGAAAAAGAACTGGCAAAAGCAGAAGTTGAAAACAAA 2901

Qy    317 GluGlnGluAla---GluLysValAsp-----------------GluGluGlu------ 327
             ||||||   |||||||||:::                       |||||||||
Db   2902 GCATTCGAGGCATTGGAAAAAGTTAACAATAATCCAAACTTGTTAGAAGAAGAGAAAAAA 2961

Qy    328 ------------------LysGluLysThrGluProAlaSerGluGluGlnGluProAla 341
                            ||||||   |||   |||   |||:::         |||
Db   2962 GCATACTTTGATGATATTAAAGAATCTAAAGAAGTTGCAGTTGAGAAAATCAATAATGCT 3021

Qy    342 GluAspThr---------------------AspGlnAlaArgLeuSerAlaAspTyrGlu 354
          |||:::|||                     |||:::|||   ::::::   :::   :::
Db   3022 GAAAATACTGCTGAAATTACGGCAGCAATTGACGAAGCGGAAATTGCATACAATGAAGAT 3081

Qy    355 LysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGluLys 374
             :::               |||   :::      ||||||      |||||||||
Db   3082 GTTATTAACGCAGCCCAACTTGATGCTTTGAATAAGCTTGAAAAAGATAGCGAAGAAACT 3141

Qy    375 CysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGluValVal 394
          |||   :::                                    :::|||:::   :::
Db   3142 AAGGCAGCTATT--------------------------GATGCTAATCCAAACTTAACT 3174

Qy    395 AlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGlu 414
             |||               :::   |||   ||||||
Db   3175 CCGGAAGAGAAAGCGAAAGCTATTGCTAAGGTAGAAGAGCTTGTT---------AATAAT 3225

Qy    415 AlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAla 434
          ||||||      ::::::   :::   |||   |||::::::
Db   3226 GCTGAATCTGACATTTTGTCGAAGCCTACCCCAGAAACAGTT------------------ 3267

Qy    435 GluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGlu 454
                |||   |||      :::|||:::      ::::::|||   |||
Db   3268 ------CAAGCAGTGGAGGATAAGGCTGAC------AAAGATCTTGCCAAAGTAGAACTT 3315

Qy    455 MetCysValSerGlyGlyAspHisThrGlnLeu---------ThrAspLeuSerProGlu 471
                :::   |||            :::            :::|||:::||||||
Db   3316 CAAGCAGCAGCAGACGGTGCGAAGAAAGGCATTGAAGCAAATCCGAATTTGACTCCAGAA 3375

Qy    472 GluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSer 491
          ||||||   :::   |||                                 :::
Db   3376 GAGAAAGATGTAGCTAAGAAG-------------------------------GCAGTA 3402

Qy    492 GlnGluArgIleLysValGlnGlySerProLeuLysLys--------------------- 504
          ::::::   :::||||||            :::   |||
Db   3403 GAAGACGCGGTTAAGGTGGCGACAGACGCTATTGATAAGGCGTCAACTCCAACCGAAGTT 3462

Qy    505 ---LeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArg 523
                :::|||   |||:::|||   :::      ::::::   |||         :::
Db   3463 GACACAGCGACAAGCGATGGAGTGAAGGCTATTGATGCAGAAGAGTTTAAAGCTACTCAG 3522

Qy    524 GlyGlyGlyGlyAsp---------GluGluProGlyGluTyrGlnHisIleHisThrGlu 540
                   :::            :::|||         :::               :::
Db   3523 AAAGATGCTAAGAACAAGATTGCCAAAGAAGCAGAATCAGCTAAGAAAGCGATTGACGAC 3582

Qy    541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
          :::|||            ::::::|||   ||||||:::   ::::::   ||||||
Db   3583 AATCCAAACTTGACTCCAGATGAGAAG---GAATCAGCTAAGAATGCAGTGGAAGAGGCG 3639

Qy    561 GluGlu------ThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGlu 578
             :::         |||      ::::::|||               |||         :::
Db   3640 GCTAAGGTAGCAACAGCCGCTATTGATAAAGCA------TCAACTCCAGATGCAGTTCAA 3693

Qy    579 AlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAla 598
             ||||||               |||               ||||||      |||
Db   3694 GTAGAAGAG---------GACAAAGGTGTAGCAGCTATCAATTTGATTACTGCCAAGGCA 3744

Qy    599 SerPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLys 618
                |||   ::::::   |||                           ||||||
Db   3745 GATGCTAAAGGTGTCATTGCTGCTAAGTTGGCAGATGAAATCAAGAAGCTCGAAGATAAG 3804

Qy    619 GluGluGluLeuGluLys---ValLysSerAlaThrLeuSerSerThrAspSerThrVal 637
          :::   |||   ||||||   :::   ::::::|||:::::::::   :::      :::
```

```
Db    3805 CAAGCAGAAGCAGAAAAAGCTATCGATGCGTCAACTATGACTAATGAGGAGAAAGCAATC 3864

Qy     638 SerGlu------MetGlnAspGluValLysThrValGlyGluGlnLysProGlu--- 654
           ::::::      :::|||||  |||              ::::::    |||   |||
Db    3865 GCTAAGAAGGCTCTTCAAGATGTTGTA------------GATAAAGGAAAAGCAGAGCTT 3912

Qy     655 GluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySer 674
           |||          ||||||  |||:::         ||||||             :::
Db    3913 GAAGACGCAGCTAGGGTAGCAACAAATGAGATTCATGAAGCT---------------ACT 3957

Qy     675 SerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeu 694
           :::   ::::::|||:::   |||      :::   ::::::              |||
Db    3958 ACTACAGAAAAAGCGAAAGCGGCGGAACTTGCTGGCGAAAAGAGCTTGACAGACACA--- 4014

Qy     695 GlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAla 714
           |||   ::::::       |||   |||    :::||||||||||||       :::|||
Db    4015 GGTAAAGAAGCTAGAGATGCAGTTGAATTGGCTAAGGATAAAGAATTAGCTAAGGAAGCA 4074

Qy     715 ValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAla 734
           :::            :::|||      :::                :::   :::
Db    4075 ATCCGAACAGAAGAAGAAGAAGCTACTAAAAATAGTAGAGAAACTTGCAGAAGATACGCGC 4134

Qy     735 GlySerProSerGluGlyGluGlyValSerThrTrpGlu---------SerPheLysArg 751
              :::       :::         :::|||   :::             |||:::
Db    4135 AAAGCTATCGAGGACAATCCAAACTTGTCAGATGAAGATAAGCAAGCGGAAATTAAAAAG 4194

Qy     752 LeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSer 771
           |||               |||:::   :::   :::   :::   |||:::      :::
Db    4195 CTAACTGACGCTGTGGCAAAAACTTTAGCAACCATTCGTGACAATGCAGATAAGCGTACG 4254

Qy     772 ValGlu-----------GlnLeuSerThrGluIleGluProSerArgGluGluSerTrp 787
              |||           |||      ::::::|||    ::::::|||
Db    4255 CAAGAAGCAGAAAAAGCTCAAGCCCTAGCAGATCTTGAAAAAGCTAAAGAAACACAGAAA 4314

Qy     788 ValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGlu 807
           ::::::      |||         |||                  ||||||:::   |||
Db    4315 ATTGCAGATAAAGCTGCGATTGATAGGTTGACTATACTTGTGAAAGATGGTGAGCTTGAA 4374

Qy     808 GlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValPro 827
              ||||||   :::|||
Db    4375 ---GCTACTAAACAAGAT------------------------------------------ 4389

Qy     828 AlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGly 847
           |||            :::::::::         |||      :::|||   :::   :::
Db    4390 GCTAAGAACAAGATTGCTAAAGATGCAGCCGCTGCTAAAGAAGCAATTGCAAGCAATCCA 4449

Qy     848 AsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThr 867
           |||                                 :::::::::          ||||||
Db    4450 AAC--------------------------------TTGACAGACGCAGAGAAGAAAACC 4476

Qy     868 LeuValHisThrValSerValAlaValIle---------AspGlyThrArgAlaValThr 884
                  |||      |||                     |||         |||   |||
Db    4477 TTCACCGATGCGGTAGATGCAGAAGTAGCCAAAGCTAACGACGCAATTTCAGCTGCAACC 4536

Qy     885 Ser---------ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluPro 901
           |||         |||::::::           :::   ::::::|||         :::
Db    4537 AGCCCAGCAGATGTTCAAAAAGAAGAGGATGCAGGTGTTGCAGCCATTGCAGAAGATGTT 4596

Qy     902 LeuGluHisThrAlaGlyGluAlaMetProProVal------GluGluValThrGluLys 919
           |||:::               :::|||      :::   :::               |||
Db    4597 CTTGACGCAGCTAAACAAGATGCTAAGAATAAGATTGCTAAAGATGCAGCCGCTGCTAAA 4656

Qy     920 AspIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAla 939
           :::     |||             |||   ||||||            |||
Db    4657 GAAGCAATTGGC---TCCAATCCAAACTTGACAGACGCA---------GAGAAGAAAACC 4704

Qy     940 HisAspAspMetValThrSerGluValAspPheThrSerGluAlaValThr---AlaThr 958
              |||   |||   :::||||||      ::::::|||::::::      ||||||
Db    4705 TTCACCGATGCGGTAGATGCAGAAGTAGCCAAAGCTAACGACGCAATTTCAGCTGCAACC 4764

Qy     959 GluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThr 978
                 ::::::      :::   |||:::         |||:::   ||||||:::
Db    4765 AGCCCAGCAGATGTTCAAAAAGAAGAGGATGCAGGTGTTGCAGCCATTGCAGAAGATGTT 4824

Qy     979 ThrAspMetVal---------------SerAlaValSerGlnLeuThrAspSerProAsp 993
              |||            :::  :::::::::   :::|||:::
Db    4825 CTTGACGCAGCTAAACAAGATGCTAAGAATAAGATTGCTAAAGAATCCGACGCTGCTAAG 4884

Qy     994 ThrThrGluGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGlu 1013
           :::      :::|||   |||                  :::   |||   |||:::
Db    4885 TCAGCCATTGACGCGAATCCA------------------AACTTGACAGATGCAGAGAAG 4926

Qy    1014 Glu------GluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGlu 1031
           |||      ::::::            |||      :::|||      |||   :::
Db    4927 GAATCAGCTAAGAAAGCAGTTGATGCTGATGCTAAAGCTGCGACAGATGCAATTGATGCT 4986

Qy    1032 GluSerGlnValProAlaThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLys 1051
              :::                  |||:::   :::      |||
Db    4987 TCAACAAGTCCAGTCGAAGCGCAATCGGCAGAGGACAAAGGC------------------ 5028

Qy    1052 ValGluGluValGluGluAspSerGluValLeuAlaSerGluLysGluLysAspValMet 1071
           |||      :::   :::|||                |||   :::      |||:::   :::
Db    5029 GTAGGTTCAATCGCCCAAGATGTTCTTGACGCAGCGAAACAAGATGCTAAGAACAAGATT 5088

Qy    1072 ProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGly 1091
                                              |||:::   :::|||      |||
Db    5089 --------------------------GCCAAAGAAGTTGCCGCAGCTAAAGAAGCAATT 5121

Qy    1092 GlnAlaThrProGlu-----------SerLeuGluValProGluValThrAlaAspVal 1107
              |||   |||                     |||      :::         |||
Db    5122 GATGCCAATCCGAACTTATCAGATGCAGAGAAGGAAGCTTCTAAGAAAGCGGTAGATGCA 5181

Qy    1108 AspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnAlaValAla 1127
              |||   ||||||         |||                  :::   :::   :::
Db    5182 GATGCTAAAGCTACGACAGATGCAATT------------------GATGCTTCAACAAGT 5223

Qy    1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
              |||   :::::::::         :::|||      |||         :::|||
Db    5224 CCAGTCGAAGCGCAATCG---------GCAGAGGACAAAGGCGTAGGTTCAATCGCC--- 5271

Qy    1148 SerAspThrAlaAspGlyThrGlnGlnAsp---------GluThrIleAspSerGlnAsp 1164
              |||      |||      :::||||||         :::               |||
Db    5272 CAAGATGTTCTTGACGCAGCGAAACAAGATGCTAAGAATAAGATTGCTAAAGAATCCGAC 5331

Qy    1165 SerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAla 1184
           :::      :::|||               :::|||:::   |||      :::|||
Db    5332 GCTGCTAAGTCAGCCATTGACGCGAATCCAAACTTGACAGATGCAGAGAAGGAATCAGCT 5391

Qy    1185 GlnLysGlu-----------------------------------GluProSerThrLeu 1192
           :::|||                                      :::   ||||||
Db    5392 AAGAAAGCGGTAGATGCAGATGCTAAAGCTGCGACAGATGCAATTGATGCTTCAACAAGT 5451

Qy    1193 ProAsnAsnValProAlaGlnGluGluHisGlyGluGluPro---GlyArgAspValLeu 1211
              |||            :::   |||:::   |||         :::|||:::|||
Db    5452 CCAGTCGAAGCGCAATCGGCAGAGGACAAAGGCGTAGGCGCCATCGCCAAAGACATTCTT 5511

Qy    1212 GluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluVal 1231
           :::   :::|||:::                              ||||||   ::::::
Db    5512 GATGCCGCGAAACAAGAT--------------------------GCTAAGAACAAGATT 5544

Qy    1232 GlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGluGluGlnGluVal 1251
              :::|||                                    |||      :::
Db    5545 GCTAAAGAGGCA------------------------------GAATCCGCTAAGTCA 5571

Qy    1252 PheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGlu--- 1270
              :::   |||   ||||||         ||||||               ||||||
Db    5572 GTCATTGACTCCAATCCGAACTTGACAGATGCAGCTAAGGAAGCGGCTAAATCTGAAATT 5631

Qy    1271 ------------------------------ValMetGlyValAlaGlyCysGlnGluLys 1280
                                         :::   ||||||         ||||||
Db    5632 GATAAAGCTGTTGAGGAAGCGATTGTTTTAATCAATGGTGTTAGAACTTATCAAGAGTTG 5691

Qy    1281 GluSerThrGluValGlnSerLeuSerLeu------------------------------ 1290
           |||   ::::::         :::|||
Db    5692 GAAAAAATCAAACTTCCAATGGCAGCTCTAATTAAACCAGCTGCGAAAGTAACACCAGTG 5751
```

```
Qy   1291 ------GluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysPro 1308
                                 |||    ::::::  :::    |||         |||
Db   5752 GTTGATCCAAATAACTTGACTGAAAAAGAAATTGCTCGTATCAAGGCATTCCTTAAAGAG 5811

Qy   1309 GluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyr 1328
                            |||     :::          :::   :::          |||
Db   5812 AACAATAACCTCCCATAAGGAACAGAGATTAATGTTTCTAAAGATGCTTCAGTGACAATT 5871

Qy   1329 GlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySer 1348
                |||       |||:::|||:::   |||
Db   5872 AAATATCCAGATGGAACTATTGATTTGCTATCACCA------------------------ 5907

Qy   1349 LeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnVal 1368
                                                      :::||||||    |||
Db   5908 ---------------------------------------------GTAGAAGTT---GTG 5919

Qy   1369 GlnSerLeuAspThrThrValThrGlnThrAlaGlu------------------------ 1380
          :::       |||   |||            |||
Db   5920 AAGCAGGCAGATAAAACTGCTCCTACGGTCGCAAATGATGGCAAAGGTAATATTGTGATT 5979

Qy   1381 ---AlaValGluLysValIleGluThrValVal------IleSerGluThrGlyGluSer 1397
                   ||||||   :::|||   ||||||       :::          |||:::|||
Db   5980 GTACCGTCTGAAAAAGCTGTTGAGCTTGTTGTTTCATACGTAGATAACAATGGTAAGTCG 6039

Qy   1398 ProGluCysValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHis 1417
                   |||                                 :::    |||   |||
Db   6040 CAAACTGTAGTT-----------------------GTTACGAAAGGTACGGATGGTTTA 6075

Qy   1418 TrpThrLeuGlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSer 1437
          ||||||           |||   :::||||||
Db   6076 TGGACA---------GCAAGTAATACAGTG------------------------------ 6096

Qy   1438 IleProIleIleValThrPro------------------AlaProGluSerThrLeuHis 1451
                :::||||||   |||                   |||   |||    :::
Db   6097 ------GTGATTGTGGACCCTGTGACTGGGCAAGTAATCGTTCCAGGTTCTGTTATTAAG 6150

Qy   1452 ProAspLeuGlnGlyGluIle------SerAlaSerGlnArgGluArgSerGluGluGlu 1469
          |||         |||   :::     :::   |||:::   |||
Db   6151 CCA---------GGTACAGTTGTTACAGCATACTCTAAAGACGAGGTTGGAAATAGTTCT 6201

Qy   1470 AspLysProAspAlaGlyProAspAla---AspGlyLysGluSerThrAlaIleGluLys 1488
          |||      :::|||         |||   |||          |||   |||      |||
Db   6202 GATTCAGCAGAAGCTGAAGTTGTAGCAGTAGACGAAAATAATTCTGCAGCAGGAGTGAAA 6261

Qy   1489 ValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIle 1506
          |||                             :::|||   |||:::   ::::::
Db   6262 GTTAAATCAGTTACTACAAATGCTAATAATGTTGAGAAGAAAGCTAAGCAATTA 6315
```

RESULT 7
US-08-466-390-3
; Sequence 3, Application US/08466390
; Patent No. 5686562
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA, HURWITZ & THIBEAULT
; STREET: 125 HIGH STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,390
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PITCHER ESQ, EDMUND R
; REGISTRATION NUMBER: 27,829
; REFERENCE/DOCKET NUMBER: MTP-013
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 248-7000
; TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6306 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..6306
; PUBLICATION INFORMATION:
; AUTHORS: COMPTON, DUANE A
; AUTHORS: SZILAK, ILYA
; AUTHORS: CLEVELAND, DON W
; TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
; TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
; TITLE: SEGREGATION OF PROTEINS AT MITOSIS
; JOURNAL: J. Cell Biol.
; VOLUME: 116
; PAGES: 1395-1408
; DATE: 1992
US-08-466-390-3

Alignment Scores:

Pred. No.:	1.98e-19	Length:	6306
Score:	415.00	Matches:	373
Percent Similarity:	35.28%	Conservative:	255
Best Local Similarity:	20.96%	Mismatches:	663
Query Match:	5.14%	Indels:	493
DB:	1	Gaps:	72

US-09-902-432-4 (1-1596) x US-08-466-390-3 (1-6306)

```
Qy     28 GlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspPro 47
          ||||||         |||         :::|||   ||||||:::   |||         |||
Db   1050 GGCCACTCAGGAGTGGCTAGAGAAGCAGGCCCAGCTGGAGAAGGAGCTCAGCGCAGCCCT 1109

Qy     48 Ala-ThrLys------LeuProGlnLysAsn-----------GlyGlnLeuSerSerVa 61
          |||   |||       |||   :::||||||           |||:::||||||  ::
Db   1110 GCAGGACAAGAAATGCCTTGAAGAGAAGAACGAAATCCTTCAGGGAAAACTTTCACAGCT 1169

Qy     61 l--------------------------AsnGlyValAlaGluGlnGlyAspVal----- 70
          :                           |||      |||:::|||:::|||
Db   1170 GGAAGAACACTTGTCCCAGCTGCAGGATAACCCACCCCAGGAGAAGGGCGAGGTGCTGGG 1229

Qy     71 ---------HisValGlnGluGluAsnGlnGluGly----------------------- 79
                   ::::::      ||||||
Db   1230 TGATGTCTTGCAGCTGGAAACCTTGAAGCAAGAGGCAGCCACTCTTGCTGCAAACAACAC 1289

Qy     80 ---------GlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAs 96
                            |||:::::::::   :::   ||||||:::|||::::::
Db   1290 ACAGCTCCAAGCCAGGGTAGAGATGCTGGAGACTGAGCGAGGCCAGCAGGAAGCCAAGCT 1349

Qy     96 pValArgGluLysAspArgValGluGlu-------------------------------- 105
           :::   |||:::          ||||||
Db   1350 GCTTGCTGAGCGGGGCCACTTCGAAGAAGAAAAGCAGCAGCTGTCTAGCCTGATCACTGA 1409

Qy    106 -------------MetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGl 121
                       ::::::      |||   |||   |||::::::    :::      ||
Db   1410 CCTGCAGAGCTCCATCTCCAACCTCAGCCAGGCCAAGGAAGAGCTGGAGCAGGCCTCCCA 1469

Qy    121 nGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMe 141
          |            :::   :::      |||:::   :::      :::   :::     ::
Db   1470 GGCTCATGGGGCCCGGTTGACTGCCCAGGTGGCCTCTCTGACCTCTGAGCTCACCACACT 1529
```

```
Qy    141 tValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheVa 161
          :            :::   |||      ::::::
Db   1530 CAATGCCACCATCCAGCAACAGGATCAAGAACTG-----------------------GC 1565

Qy    161 lGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLe 181
           |||    |||          |||:::   :::   :::   :::   |||:::|||
Db   1566 TGGCCTGAAGCAGCAGGCCAAAGAGAAGCAGGCCCAGCTAGCACAGACCCTCCAA----- 1620

Qy    181 uThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGl 201
                                  :::     :::||||||   |||       |||||
Db   1621 ----------------------CAGCAAGAACAGGCCTCCCAGGGCCTC---CGCCACCA 1655

Qy    201 nGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGl 221
          |         ||||||                     ||||||   :::||||||::
Db   1656 G---------GTGGAGCAGCTAAGCAGTAGCCTGAAGCAGAAGGAGCAGCAGTTGAAGGA 1706

Qy    221 nSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLe 241
          :        ||||||||||||   |||      ::::::      :::   ::::::
Db   1707 GGTAGCGGAGAAGCAGGAGGCAACT------AGGCAGGACCATGCCCAGCAACTGGCCAC 1760

Qy    241 uGlnAlaGluSerAspGlnAlaAla------GluGluGluAlaLysAspGluGlyGluGl 259
             ||||||      :::|||:::      |||   :::|||      :::      ||
Db   1761 TGCTGCAGAGGAGCGAGAGGCCTCCTTAAGGGAGCGGATGCGGCTCTCAAGCAGCTGGA 1820

Qy    259 uLysGlnGluLysGluProThrLysSerProGlu------SerProSerSerProValAs 277
          |         |||||||||                |||                |||
Db   1821 GGCACTGGAGAAGGAGAAGGCTGCCAAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGC 1880

Qy    277 nSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLy 297
           :::|||              |||         |||                 |||::
Db   1881 TAATGAAGCCCGGGACAGTGCCCAGACCTCAGTGACACAG---------GCCCAGCGGGA 1931

Qy    297 sLysThrSerPheLysLysSerLysGluAsp---------AspLeuGluThrAlaGluLy 314
          :|||           :::      |||:::            :::||||||||||   ::
Db   1932 GAAGGCAGAGCTGAGCCGGAAGGTGGAGGAACTCCAGGCCTGTGTTGAGACAGCCCGCCA 1991

Qy    314 sArgLysGluGlnGluAlaGluLysValAspGluGluGlu---------LysGluLysTh 331
          :   :::   :::         :::|||   |||   |||         :::   :::
Db   1992 GGAACAGCATGAGGCCCAGGCCCAGGTTGCAGAGCTAGAGTTGCAGCTGCGGTCTGAGCA 2051

Qy    331 rGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAl 351
           :::     |||:::|||::::::         |||::::::   ||||||      |||
Db   2052 GCAAAAAGCAACTGAGAAAGAAAGGGTGGCCCAGGAGAAGGACCAG------CTCCAGGA 2105

Qy    351 aAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSe 371
           :::     ::::::         |||:::         |||   ||||||
Db   2106 GCAGCTCCAGGCCCTCAAAGAGTCCTTGAAGGTCACCAAGGGCAGCCTTGAA-------- 2157

Qy    371 rGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGl 391
           |||||||||         |||   :::      :::|||:::
Db   2158 -GAGGAGAAG------CGCAGGGCTGCAGATGCCCTGGAAGAGCAGCAGCGTTGTATCTC 2210

Qy    391 nGluValValAlaGluVal-----------------HisValSerThrValGluLysTh 405
           |||:::   ||||||                  |||               ::::::
Db   2211 TGAGCTGAAGGCAGAGACCCGAAGCCTGGTGGAGCAGCATAAGCGGGAACGAAAGGAGCT 2270

Qy    405 rGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGl 425
           |||||||||:::   |||               |||      ::::::      |||
Db   2271 GGAAGAAGAGAGGGCTGGGCGCAAGGGGCTGGAGGCTCGATTACTGCAGCTTGGGGAGGC 2330

Qy    425 yGluSerLeuProProGlu------------LysLeuAlaGluProGlnGluValProGl 441
                         |||             :::|||||||||
Db   2331 CCATCAGGCTGAGACTGAAGTCCTGCGGCGGGAGCTGGCAGAGGCCATGGCTGCCCAGCA 2390

Qy    441 nGluAlaGlu---ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGl 460
              ||||||      |||:::|||:::|||                        |||
Db   2391 CACAGCTGAGAGTGAGTGTGAGCAGCTCGTCAAAGAAGTAGCTGCCTGGCGTGACGGGTA 2450

Qy    460 yAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGl 480
           :::   :::|||                  ||||||   |||:::   :::|||
Db   2451 TGAGGATAGCCAG---------------CAAGAGGAGGC-ACAGTATGGCGCCATGTTCC 2494

Qy    480 uGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySe 500
           |||          :::   |||
Db   2495 AGGAACAGCTGATGACTTTGAA-------------------------------------- 2516

Qy    500 rProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLy 520
                                                         |||        ::
Db   2517 ---------------------------------------------GGAGGAATGTGAGA 2530

Qy    520 sGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGl 540
          :|||      |||      |||   :::      |||                  |||
Db   2531 AGGCCCGCCAGGAGCTGCAGGAGGCAAAGGAGAAGGTGGCAGGCATAGAATCCCACAGCG 2590

Qy    540 uSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPr 560
           :::|||         |||:::         |||   ::::::      :::|||   :::||
Db   2591 AGCTCCA---------GATAAGCCGGCAGCAGAACAAACTAGCTGAGCTCCATGCCAACC 2641

Qy    560 oGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAla-- 579
          |      ::::::|||            ||||||   |||      |||      ||||||
Db   2642 TGGCCAGAGCACT------CCAGCAGGTCCAAGAGAAGGAAGTCAGGGCCCAGAAGCTTG 2695

Qy    580 ------------------------GluGluGly-ThrThrSerAspGlyGluLysLysA 591
                                  ::::::|||   ||||||             :::
Db   2696 CAGATGACCTCTCCACTCTGCAGGAAAAGATGGCTGCCACCAGCAAAGAGGTGGCCCGCT 2755

Qy    591 rgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValA 611
             |||   :::         |||      ::::::            ::::::   |||:
Db   2756 TGGAGACCTTGGTCCGCAAGGCAGGTGAGCAGCAGGAAACAGCCTCCCGGGAGTTAGTCA 2815

Qy    611 rgArgProSerGluSer---AspLysGluGluGlu---LeuGluLysValLysSerAlaT 629
           ::    |||:::   :::   |||::::::   |||   ||||||:::   :::
Db   2816 AGGAGCCTGCGAGGGCAGGAGACAGACAGCCCGAGTGGCTGGAAGAGCAACAGGGACGCC 2875

Qy    629 hrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyG 649
               ||||||   :::   :::      |||:::   |||      :::   :::|||
Db   2876 AGTTCTGCAGCACACAGGCAGCGCTGCAGGCTATGGAGCGGGAGGCAGAGCAGATGGGCA 2935

Qy    649 luGlu----------------------------------------GlnLysProGluGlu- 655
             |||                                         |||:::   ||||||
Db   2936 ATGAGCTGGAACGGCTGCGGGCCGCGCTGATGGAGAGCCAGGGGCAGCAGCAGGAGGAGC 2995

Qy    655 ------------------------------------------------------------ 655

Db   2996 GTGGGCAGCAGGAAAGGGAGGTGGCGCGGCTGACCCAGGAGCGGGGCCGTGCCCAGGCTG 3055

Qy    656 -----------------------------------------------------------P 656

Db   3056 ACCTTGCCCTGGAGAAGGCGGCCAGAGCAGAGCTTGAGATGCGGCTGCAGAACGCCCTCA 3115

Qy    656 roLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerL 676
             ::::::|||||||||:::      :::                     :::      |
Db   3116 ACGAGCAGCGTGTGGAGTTCGCTACCCTGCAAGAGGCACTGGCTCATGCCCTGACGGAAA 3175

Qy    676 ysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyG 696
            ||:::                           |||   |||   |||      :::      |
Db   3176 AGGAA------------------GGCAAGGACCAGGAGTT-GGCCAAGCTTCGTGGTCTG 3216

Qy    696 lyAspSer------HisArgAlaGluGluAlaSerLysAspLysGluAla------GlyT 712
            ||   |||         ||||||         |||      ::::::||||||      ||||
Db   3217 GAGGCAGCCCAGATAAAAGAGCTGGAGGAACTTCGGCAAACCGTGAAGCAACTGAAGGAA 3276

Qy    712 hrAspAlaValProAla-SerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPro 731
            ||               |||   :::      :::|||::::::   |||      ||||||   :::
Db   3277 CA-------------GCTGGCTAAGAAAGAAAAGGAGCACGCATCTGGCTCAGGAGCCCAA 3324

Qy    732 GluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArg 751
             |||         :::|||      |||                  |||:::   :::
Db   3325 TCTGAGGCTGCTGGCAGGACAGAGCCAACAGGCCCCAAGCTG---GAAGCACTGCGGGCA 3381

Qy    752 LeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGlu--------- 768
             |||:::         ::::::         |||         :::|||:::|||:::
Db   3382 GAGGTGAGCAAGCTGGAACAGCAATGCCAGAAGCAGCAGGAGCAGGCTGACAGCCTGGAA 3441

Qy    769 -----------------------------------------AspSerSerValGluGlnLeu 775
                                                   ||||||::::::|||   |||
```

```
Db   3442 CGCAGCCTCGAGGCTGAGCGGGCCTCCCGGGCTGAGCGGGACAGTGCTCTGGAGACTCTG 3501
Qy    776 SerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIlePro 795
Db   3502 CAGGGCCAGTTAGAGGAGAAGGCCCAGGAGCTAGGGCACAGTCAGAGTGCCTTAGCCTCG 3561
Qy    796 GlyArgArgLysLysArgAla------AspGlyLysGlnGluGlnAlaThrValGluAsp 813
Db   3562 GCCCAACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAAGACCACAGCAAGGCTGAAGAT 3621
Qy    814 SerGlyProValGluIle------------AsnGluAspAspProAsnValProAlaVal 829
Db   3622 GAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCTGAGAGGAAAAATAGCCTCATC 3681
Qy    830 ValProLeu---SerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsn 848
Db   3682 AGCAGCTTGGAGGAGGAGGTGTCCATCCTGAATCGCCAGGTCCTGGAGAAGGAGGGG--- 3738
Qy    849 ThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeu 868
Db   3739 ---GAGAGCAAGGAGTTGAAGCGGCTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTG 3795
Qy    869 ValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGlu 888
Db   3796 GAGGAGAGCTGCGCCTGCTGC------------------------------AGGCAG 3822
Qy    889 ArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGlu 908
Db   3823 AGACAGCCAGCA------------------------------------------------ 3834
Qy    909 AlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVal 928
Db   3835 ---------ACAGTGCCAGAGCTGCAGAACGCAGCTCTGCTCTGCGGGAGG--------- 3876
Qy    929 LeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGlu--- 947
Db   3877 ------AGGTGCAGAGCCTCCGGGAGGGAGGCTGAGAAACAGCGGGTGGCTTCAGAGAAC 3930
Qy    948 ------ValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArg 965
Db   3931 CTGCGGCAGGAGCTGACCTCACAGGCTGAGCGTGCGGAGGAGCTGGGCCAAGAATTGAAG 3990
Qy    966 Thr------GluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSer 983
Db   3991 GCGTGGCAGGAGAAGTTCTTCCAGAAAGAGCAGGCCCTCTCCACCCTGCAGCTC------ 4044
Qy    984 AlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGlu 1003
Db   4045 ---------------------------------------GAGCACACCAGCACACAGGCC 4065
Qy   1004 ValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGln 1023
Db   4066 CTGGTGAGTGAGCTGCTGCCAGCTAAGCACCTCTGCCAGCAGCTGCAGGCCGAGCAGGCC 4125
Qy   1024 AlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArg 1043
Db   4126 GCTGCCGAGAAACGCCACCGTGAGGAGCTGGAG-----------CAGAGCAAGCAGGCC 4173
Qy   1044 ThrGlySerLysAlaLeuGlu--------------------------------------- 1050
Db   4174 GCTGGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAGCGGGAGCTTGGGGAGCTGATTCCT 4233
Qy   1051 ---------LysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys--- 1066
Db   4234 CTGCGGCAGAAGGTGGCAGAGCAGGAGCGAACAGCTCAGCAGCTGCGGGCAGAGAAGGCC 4293
Qy   1066 ------------------------------------------------------------ 1066
Db   4294 AGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTGCTGGCAGAGGAGAAC 4353
Qy   1066 ------------------------------------------------------------ 1066
Db   4354 CGGGGGCTGGGTGAGCGGGCCAACCTTGGCCGGCAGTTTCTGGAAGTGGAGTTGGACCAG 4413

Qy   1067 ------GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeu 1084
Db   4414 GCCCGGGAAAAGTATGTCCAAGAGTTGGCAGCCGTACGTGCTGATGCTGAGACCCGTCTG 4473
Qy   1085 AlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThr 1104
Db   4474 GCTGAGGTGCAGCGAGAAGCACAGAGCACTGCCCGGGAGCTGGAGGTG------ATGACT 4527
Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGln 1124
Db   4528 GCCAAGTATGAGGGTGCC-----------------AAGGTCAAGGTCCTGGAGGAGAGG 4569
Qy   1125 AlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrPro 1144
Db   4570 CAGCGGTTCCAGGAAGAGAGGCAGAAACTCACTGCCCAGGTGGAAGAACTGAGTAAGAAA 4629
Qy   1145 LeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAsp------GluThrIleAspSer 1162
Db   4630 CTGGCTGACTCTGACCAAGCCAGCAAGGTGCAGCAGCAGAAGCTGAAGGCTGTCCAGGCT 4689
Qy   1163 Gln------AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
Db   4690 CAGGGAGGCGAGAGCCAGCAGGAGGCCCAGCGCTTCCAGGCCCAGCTGAATGAACTGCAA 4749
Qy   1181 AlaAlaThrAlaGlnLysGluGluProSer-----------------------ThrLeu 1192
Db   4750 GCCCAGTTGAGCCAGAAGGAGCAGGCAGCTGAGCACTATAAGCTGCAGATGGAGAAAGCC 4809
Qy   1193 ProAsnAsnValProAlaGlnGluGluHisGlyGluGlu---------------ProGly 1207
Db   4810 AAAACACATTATGATGCCAAGAAGCAGCAGAACCAAGAGCTGCAGGAGCAGCTGCGGAGC 4869
Qy   1208 ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAla 1227
Db   4870 CTGGAGCAGCTGCAGAAGGAAAACAAAGAGCTGCGAGCTGAAGCT--------------- 4914
Qy   1228 LysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu 1247
Db   4915 ---GAACGGCTGGGCCATGAGCTACAG---------CAGGCTGGGCTGAAGACCAAGGAG 4962
Qy   1248 GluGlnGluValPheValHis------SerGlyProAsnSerGlnLysAlaAlaAspVal 1265
Db   4963 GCTGAACAGACCTGCCGCCACCTTACTGCCCAGGTGCGCAGCCTGGAGGCACAGGTTGCC 5022
Qy   1266 ThrTyrAspSerGluValMetGlyValAlaGlyCysGln------------GluLysGlu 1281
Db   5023 CATGCAGACCAGCAGCTTCGAGACCTGGGCAAATTCCAGGTGGCAACTGATGCTTTAAAG 5082
Qy   1282 SerThrGluValGlnSer-----------------------------------LeuSer 1289
Db   5083 AGCCGTGAGCCCCAGGCTAAGCCCCAGCTGGACTTGAGTATTGACAGCCTGGATCTGAGC 5142
Qy   1290 LeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGlu 1309
Db   5143 TGCGAGGAGGGGACCCCACTCAGTATCACCAGCAAGCTGCCTCGT---ACCCAGCCAGAC 5199
Qy   1310 GlnValSerGluGluGlyGluGlnGlu---------ThrAlaAlaProGluHisGluGly 1326
Db   5200 GGCACCAGCGTCCCTGGAGAACCAGCCTCACCTATCTCCCA-GCGCCTGCCCCCCAAGGT 5258
Qy   1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeu 1346
Db   5259 AGAATCCCTGGAGAGTCTCTACTTCACTCCCATCCCTGCTCGGAGTCAGGCCCCCCT--- 5315
Qy   1347 GlySerLeuGlyGlySerProSerLeuPro----------------------------- 1356
Db   5316 ---------GGAGAGCAGCCTGGACTCCCTGGGAGACGTCTTCCTGGACTCGGGTCGTAA 5366
Qy   1357 ------------------------------------------------AspGlnAspLys 1360
Db   5367 GACCCGCTCCGCTCGTCGGCGCACCACGCAGATCATCAACATCACCATGACCAAGAA--- 5423
```

```
Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeu-AspThrThrValThrGlnThrAlaGl 1380
          |||   |||         :::   :::   |||             |||:::::|||
Db   5424 GCTAGATGTGGAAGAGCCAGACAGCGCCAACTCATCGTTCTACAGCACGGTCTGCTCC 5483

Qy   1380 uAlaValGluLysValIle------------GluThrValValIleSerGluThrGlyGl 1396
           |||   :::      :::               |||      ::::::      |||
Db   5484 TGCTTCCCAGGCTAGCCTGCGAGCCACCTCCTCTACTCAGTCTCTAGCTCGCCTGGGT-- 5541

Qy   1396 uSerProGluCysValGlyAlaHisLeuLeu---------------------------- 1406
           |||||:::          :::   ||||||
Db   5542 -TCTCCCGATTATGGCAACTCAGCCCTGCTCAGCTTGCCTGGCTACCGCCCCACCACTCG 5600

Qy   1407 ----ProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAs 1425
                 |||   :::|||   |||   |||      |||               |||
Db   5601 CAGTTCTGCTCGTCGTTCCCAGGCCGG-GGTGTCCAGTGGGGCCCCTCCAGGAAGGAACA 5659

Qy   1425 pThrValPro------------------------------------LeuGlyProGluSe 1433
              :::                                       |||   |||
Db   5660 GCTTCTACATGGGCACTTGCCAGGATGAGCCTGAGCAGCTGGATGACTGGAACCGCATTG 5719

Qy   1433 rGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisProAs 1453
           :::|||   :::            |||   |||      |||      |||   |||
Db   5720 CAGAGCTGCAGCAGCGCAATCGAGTGTGCCCCCCACATCTGAAGACCTGCTATCCCCTGG 5779

Qy   1453 pLeuGlnGlyGluIleSerAlaSerGlnArgGluArg---SerGluGluGluAspLysPr 1472
           :::|||                           |||      :::|||   :::||
Db   5780 AGTCCAGGCCTTCCCTGAGCCTGGGCACCATCACAGATGAGGAGATGAAAACTGGAGACC 5839

Qy   1472 oAsp--------------------------AlaGlyProAspAlaAspGly 1480
          |                              |||   ||||||:::   |||
Db   5840 CCCAAGAGACCCTGCGCCGAGCCAGCATGCAGCCAATCCAGATAGCCGAGGG 5891
```

RESULT 8
US-08-470-950-3
; Sequence 3, Application US/08470950
; Patent No. 5698439
; GENERAL INFORMATION:
;   APPLICANT: TOUKATLY, GARY
;   APPLICANT: LIDGARD, GRAHAM P
;   TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;   TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
;   NUMBER OF SEQUENCES: 6
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: TESTA, HURWITZ & THIBEAULT
;   STREET: 125 HIGH STREET
;   CITY: BOSTON
;   STATE: MA
;   COUNTRY: USA
;   ZIP: 02110
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/470,950
;   FILING DATE: 06-JUN-1995
;   CLASSIFICATION: 536
;   ATTORNEY/AGENT INFORMATION:
;   NAME: PITCHER ESQ, EDMUND R
;   REGISTRATION NUMBER: 27,829
;   REFERENCE/DOCKET NUMBER: MTP-013
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (617) 248-7000
;   TELEFAX: (617) 248-7100
; INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 6306 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;   MOLECULE TYPE: cDNA
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: 1..6306
;   PUBLICATION INFORMATION:
;   AUTHORS: COMPTON, DUANE A
;   AUTHORS: SZILAK, ILYA
;   AUTHORS: CLEVELAND, DON W
;   TITLE: PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
;   TITLE: PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
;   TITLE: SEGREGATION OF PROTEINS AT MITOSIS
;   JOURNAL: J. Cell Biol.
;   VOLUME: 116
;   PAGES: 1395-1408
;   DATE: 1992
US-08-470-950-3

Alignment Scores:

Pred. No.:	1.98e-19	Length:	6306
Score:	415.00	Matches:	373
Percent Similarity:	35.28%	Conservative:	255
Best Local Similarity:	20.96%	Mismatches:	663
Query Match:	5.14%	Indels:	493
DB:	1	Gaps:	72

US-09-902-432-4 (1-1596) x US-08-470-950-3 (1-6306)

```
Qy     28 GlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspPro 47
           |||||         |||      :::|||   ||||||:::  |||        |||
Db   1050 GGCCACTCAGGAGTGGCTAGAGAAGCAGGCCCAGCTGGAGAAGGAGCTCAGCGCAGCCCT 1109

Qy     48 Ala-ThrLys------LeuProGlnLysAsn------------GlyGlnLeuSerSerVa 61
          |||   |||       |||   :::||||||            |||:::||||||  ::
Db   1110 GCAGGACAAGAAATGCCTTGAAGAGAAGAACGAAATCCTTCAGGGAAAACTTTCACAGCT 1169

Qy     61 l--------------------------AsnGlyValAlaGluGlnGlyAspVal----- 70
          :                           |||         |||:::|||:::|||
Db   1170 GGAAGAACACTTGTCCCAGCTGCAGGATAACCCACCCCAGGAGAAGGGCGAGGTGCTGGG 1229

Qy     71 ---------HisValGlnGluGluAsnGlnGluGly--------------------- 79
                       ::::::         ||||||
Db   1230 TGATGTCTTGCAGCTGGAAACCTTGAAGCAAGAGGCAGCCACTCTTGCTGCAAACAACAC 1289

Qy     80 ---------GlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAs 96
                             |||:::::::::   :::   ||||||:::|||::::::
Db   1290 ACAGCTCCAAGCCAGGGTAGAGATGCTGGAGACTGAGCGAGGCCAGCAGGAAGCCAAGCT 1349

Qy     96 pValArgGluLysAspArgValGluGlu------------------------------ 105
           :::   |||:::         ||||||
Db   1350 GCTTGCTGAGCGGGGCCACTTCGAAGAAGAAAAGCAGCAGCTGTCTAGCCTGATCACTGA 1409

Qy    106 ------------MetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGl 121
                       ::::::      |||   |||   |||::::::   :::      ||
Db   1410 CCTGCAGAGCTCCATCTCCAACCTCAGCCAGGCCAAGGAAGAGCTGGAGCAGGCCTCCCA 1469

Qy    121 nGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMe 141
          |         :::   :::      |||:::      :::   :::            ::
Db   1470 GGCTCATGGGGCCCGGTTGACTGCCCAGGTGGCCTCTCTGACCTCTGAGCTCACCACACT 1529

Qy    141 tValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheVa 161
          :            :::   |||      ::::::
Db   1530 CAATGCCACCATCCAGCAACAGGATCAAGAACTG-----------------------GC 1565

Qy    161 lGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLe 181
           |||   |||         |||:::   :::   :::   :::   |||:::|||
Db   1566 TGGCCTGAAGCAGCAGGCCAAAGAGAAGCAGGCCCAGCTAGCACAGACCCTCCAA----- 1620

Qy    181 uThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGl 201
                                :::      :::||||||   |||         |||||
Db   1621 ---------------------CAGCAAGAACAGGCCTCCCAGGGCCTC---CGCCACCA 1655

Qy    201 nGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGl 221
          |         ||||||                        ||||||   :::||||||::
Db   1656 G---------GTGGAGCAGCTAAGCAGTAGCCTGAAGCAGAAGGAGCAGCAGTTGAAGGA 1706
```

```
Qy    221 nSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLe 241
Db   1707 GGTAGCGGAGAAGCAGGAGGCAACT------AGGCAGGACCATGCCCAGCAACTGGCCAC 1760
Qy    241 uGlnAlaGluSerAspGlnAlaAla------GluGluGluAlaLysAspGluGlyGluGl 259
Db   1761 TGCTGCAGAGGAGCGAGAGGCCTCCTTAAGGGAGCGGGATGCGGCTCTCAAGCAGCTGGA 1820
Qy    259 uLysGlnGluLysGluProThrLysSerProGlu------SerProSerSerProValAs 277
Db   1821 GGCACTGGAGAAGGAGAAGGCTGCCAAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGC 1880
Qy    277 nSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLy 297
Db   1881 TAATGAAGCCCGGGACAGTGCCCAGACCTCAGTGACACAG---------GCCCAGCGGGA 1931
Qy    297 sLysThrSerPheLysLysSerLysGluAsp---------AspLeuGluThrAlaGluLy 314
Db   1932 GAAGGCAGAGCTGAGCCGGAAGGTGGAGGAACTCCAGGCCTGTGTTGAGACAGCCCGCCA 1991
Qy    314 sArgLysGluGlnGluAlaGluLysValAspGluGluGlu---------LysGluLysTh 331
Db   1992 GGAACAGCATGAGGCCCAGGCCCAGGTTGCAGAGCTAGAGTTGCAGCTGCGGTCTGAGCA 2051
Qy    331 rGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAl 351
Db   2052 GCAAAAAGCAACTGAGAAAGAAAGGGTGGCCCAGGAGAAGGACCAG------CTCCAGGA 2105
Qy    351 aAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSe 371
Db   2106 GCAGCTCCAGGCCCTCAAAGAGTCCTTGAAGGTCACCAAGGGCAGCCTTGAA-------- 2157
Qy    371 rGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGl 391
Db   2158 -GAGGAGAAG------CGCAGGGCTGCAGATGCCCTGGAAGAGCAGCAGCGTTGTATCTC 2210
Qy    391 nGluValValAlaGluVal------------------HisValSerThrValGluLysTh 405
Db   2211 TGAGCTGAAGGCAGAGACCCGAAGCCTGGTGGAGCAGCATAAGCGGGAACGAAAGGAGCT 2270
Qy    405 rGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGl 425
Db   2271 GGAAGAAGAGAGGGCTGGGCGCAAGGGGCTGGAGGCTCGATTACTGCAGCTTGGGGAGGC 2330
Qy    425 yGluSerLeuProProGlu------------LysLeuAlaGluProGlnGluValProGl 441
Db   2331 CCATCAGGCTGAGACTGAAGTCCTGCGGCGGGAGCTGGCAGAGGCCATGGCTGCCCAGCA 2390
Qy    441 nGluAlaGlu---ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGl 460
Db   2391 CACAGCTGAGAGTGAGTGTGAGCAGCTCGTCAAAGAAGTAGCTGCCTGGCGTGACGGGTA 2450
Qy    460 yAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGl 480
Db   2451 TGAGGATAGCCAG---------------CAAGAGGAGGC-ACAGTATGGCGCCATGTTCC 2494
Qy    480 uGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySe 500
Db   2495 AGGAACAGCTGATGACTTTGAA-------------------------------------- 2516
Qy    500 rProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLy 520
Db   2517 ---------------------------------------------GGAGGAATGTGAGA 2530
Qy    520 sGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGl 540
Db   2531 AGGCCCGCCAGGAGCTGCAGGAGGCAAAGGAGAAGGTGGCAGGCATAGAATCCCACAGCG 2590
Qy    540 uSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPr 560
Db   2591 AGCTCCA---------GATAAGCCGGCAGCAGAACAAACTAGCTGAGCTCCATGCCAACC 2641
Qy    560 oGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAla-- 579
Db   2642 TGGCCAGAGCACT------CCAGCAGGTCCAAGAGAAGGAAGTCAGGGCCCAGAAGCTTG 2695
Qy    580 ------------------------GluGluGly-ThrThrSerAspGlyGluLysLysA 591
Db   2696 CAGATGACCTCTCCACTCTGCAGGAAAAGATGGCTGCCACCAGCAAAGAGGTGGCCCGCT 2755
Qy    591 rgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValA 611
Db   2756 TGGAGACCTTGGTCCGCAAGGCAGGTGAGCAGCAGGAAACAGCCTCCCGGGAGTTAGTCA 2815
Qy    611 rgArgProSerGluSer---AspLysGluGluGlu---LeuGluLysValLysSerAlaT 629
Db   2816 AGGAGCCTGCGAGGGCAGGAGACAGACAGCCCGAGTGGCTGGAAGAGCAACAGGGACGCC 2875
Qy    629 hrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyG 649
Db   2876 AGTTCTGCAGCACACAGGCAGCGCTGCAGGCTATGGAGCGGGAGGCAGAGCAGATGGGCA 2935
Qy    649 luGlu-------------------------------------GlnLysProGluGlu- 655
Db   2936 ATGAGCTGGAACGGCTGCGGGCCGCGCTGATGGAGAGCCAGGGGCAGCAGCAGGAGGAGC 2995
Qy    655 ------------------------------------------------------------ 655
Db   2996 GTGGGCAGCAGGAAAGGGAGGTGGCGCGGCTGACCCAGGAGCGGGGCCGTGCCCAGGCTG 3055
Qy    656 -----------------------------------------------------------P 656
Db   3056 ACCTTGCCCTGGAGAAGGCGGCCAGAGCAGAGCTTGAGATGCGGCTGCAGAACGCCCTCA 3115
Qy    656 roLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerL 676
Db   3116 ACGAGCAGCGTGTGGAGTTCGCTACCCTGCAAGAGGCACTGGCTCATGCCCTGACGGAAA 3175
Qy    676 ysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyG 696
Db   3176 AGGAA-----------------GGCAAGGACCAGGAGTT-GGCCAAGCTTCGTGGTCTG 3216
Qy    696 lyAspSer------HisArgAlaGluGluAlaSerLysAspLysGluAla------GlyT 712
Db   3217 GAGGCAGCCCAGATAAAAGAGCTGGAGGAACTTCGGCAAACCGTGAAGCAACTGAAGGAA 3276
Qy    712 hrAspAlaValProAla-SerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPro 731
Db   3277 CA-----------GCTGGCTAAGAAAGAAAAGGAGCACGCATCTGGCTCAGGAGCCCAA 3324
Qy    732 GluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArg 751
Db   3325 TCTGAGGCTGCTGGCAGGACAGAGCCAACAGGCCCCAAGCTG---GAAGCACTGCGGGCA 3381
Qy    752 LeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGlu--------- 768
Db   3382 GAGGTGAGCAAGCTGGAACAGCAATGCCAGAAGCAGCAGGAGCAGGCTGACAGCCTGGAA 3441
Qy    769 --------------------------------------AspSerSerValGluGlnLeu 775
Db   3442 CGCAGCCTCGAGGCTGAGCGGGCCTCCCGGGCTGAGCGGGACAGTGCTCTGGAGACTCTG 3501
Qy    776 SerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIlePro 795
Db   3502 CAGGGCCAGTTAGAGGAGAAGGCCCAGGAGCTAGGGCACAGTCAGAGTGCCTTAGCCTCG 3561
Qy    796 GlyArgArgLysLysArgAla------AspGlyLysGlnGluGlnAlaThrValGluAsp 813
Db   3562 GCCCAACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAAGACCACAGCAAGGCTGAAGAT 3621
Qy    814 SerGlyProValGluIle------------AsnGluAspAspProAsnValProAlaVal 829
Db   3622 GAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCTGAGAGGAAAAATAGCCTCATC 3681
Qy    830 ValProLeu---SerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsn 848
```

```
Db   3682 AGCAGCTTGGAGGAGGAGGTGTCCATCCTGAATCGCCAGGTCCTGGAGAAGGAGGGG--- 3738

Qy    849 ThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeu 868
Db   3739 ---GAGAGCAAGGAGTTGAAGCGGCTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTG 3795

Qy    869 ValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGlu 888
Db   3796 GAGGAGAGCTGCGCCTGCTGC-------------------------------AGGCAG 3822

Qy    889 ArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGlu 908
Db   3823 AGACAGCCAGCA------------------------------------------------ 3834

Qy    909 AlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVal 928
Db   3835 ---------ACAGTGCCAGAGCTGCAGAACGCAGCTCTGCTCTGCGGGAGG--------- 3876

Qy    929 LeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGlu--- 947
Db   3877 ------AGGTGCAGAGCCTCCGGGAGGGAGGCTGAGAAACAGCGGGTGGCTTCAGAGAAC 3930

Qy    948 ------ValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArg 965
Db   3931 CTGCGGCAGGAGCTGACCTCACAGGCTGAGCGTGCGGAGGAGCTGGGCCAAGAATTGAAG 3990

Qy    966 Thr------GluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSer 983
Db   3991 GCGTGGCAGGAGAAGTTCTTCCAGAAAGAGCAGGCCCTCTCCACCCTGCAGCTC------ 4044

Qy    984 AlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGlu 1003
Db   4045 ---------------------------------------GAGCACACCAGCACACAGGCC 4065

Qy   1004 ValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGln 1023
Db   4066 CTGGTGAGTGAGCTGCTGCCAGCTAAGCACCTCTGCCAGCAGCTGCAGGCCGAGCAGGCC 4125

Qy   1024 AlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArg 1043
Db   4126 GCTGCCGAGAAACGCCACCGTGAGGAGCTGGAG------------CAGAGCAAGCAGGCC 4173

Qy   1044 ThrGlySerLysAlaLeuGlu--------------------------------------- 1050
Db   4174 GCTGGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAGCGGGAGCTTGGGGAGCTGATTCCT 4233

Qy   1051 ---------LysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys--- 1066
Db   4234 CTGCGGCAGAAGGTGGCAGAGCAGGAGCGAACAGCTCAGCAGCTGCGGGCAGAGAAGGCC 4293

Qy   1066 ------------------------------------------------------------ 1066
Db   4294 AGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTGCTGGCAGAGGAGAAC 4353

Qy   1066 ------------------------------------------------------------ 1066
Db   4354 CGGGGGCTGGGTGAGCGGGCCAACCTTGGCCGGCAGTTTCTGGAAGTGGAGTTGGACCAG 4413

Qy   1067 ------GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeu 1084
Db   4414 GCCCGGGAAAAGTATGTCCAAGAGTTGGCAGCCGTACGTGCTGATGCTGAGACCCGTCTG 4473

Qy   1085 AlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThr 1104
Db   4474 GCTGAGGTGCAGCGAGAAGCACAGAGCACTGCCCGGGAGCTGGAGGTG------ATGACT 4527

Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGln 1124
Db   4528 GCCAAGTATGAGGGTGCC-----------------AAGGTCAAGGTCCTGGAGGAGAGG 4569

Qy   1125 AlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrPro 1144
Db   4570 CAGCGGTTCCAGGAAGAGAGGCAGAAACTCACTGCCCAGGTGGAAGAACTGAGTAAGAAA 4629

Qy   1145 LeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAsp------GluThrIleAspSer 1162
Db   4630 CTGGCTGACTCTGACCAAGCCAGCAAGGTGCAGCAGCAGAAGCTGAAGGCTGTCCAGGCT 4689

Qy   1163 Gln------AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
Db   4690 CAGGGAGGCGAGAGCCAGCAGGAGGCCCAGCGCTTCCAGGCCCAGCTGAATGAACTGCAA 4749

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSer-----------------------ThrLeu 1192
Db   4750 GCCCAGTTGAGCCAGAAGGAGCAGGCAGCTGAGCACTATAAGCTGCAGATGGAGAAAGCC 4809

Qy   1193 ProAsnAsnValProAlaGlnGluGluHisGlyGluGlu---------------ProGly 1207
Db   4810 AAAACACATTATGATGCCAAGAAGCAGCAGAACCAAGAGCTGCAGGAGCAGCTGCGGAGC 4869

Qy   1208 ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAla 1227
Db   4870 CTGGAGCAGCTGCAGAAGGAAAACAAAGAGCTGCGAGCTGAAGCT--------------- 4914

Qy   1228 LysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu 1247
Db   4915 ---GAACGGCTGGGCCATGAGCTACAG---------CAGGCTGGGCTGAAGACCAAGGAG 4962

Qy   1248 GluGlnGluValPheValHis------SerGlyProAsnSerGlnLysAlaAlaAspVal 1265
Db   4963 GCTGAACAGACCTGCCGCCACCTTACTGCCCAGGTGCGCAGCCTGGAGGCACAGGTTGCC 5022

Qy   1266 ThrTyrAspSerGluValMetGlyValAlaGlyCysGln------------GluLysGlu 1281
Db   5023 CATGCAGACCAGCAGCTTCGAGACCTGGGCAAATTCCAGGTGGCAACTGATGCTTTAAAG 5082

Qy   1282 SerThrGluValGlnSer------------------------------------LeuSer 1289
Db   5083 AGCCGTGAGCCCCAGGCTAAGCCCCAGCTGGACTTGAGTATTGACAGCCTGGATCTGAGC 5142

Qy   1290 LeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGlu 1309
Db   5143 TGCGAGGAGGGGACCCCACTCAGTATCACCAGCAAGCTGCCTCGT---ACCCAGCCAGAC 5199

Qy   1310 GlnValSerGluGluGlyGluGlnGlu---------ThrAlaAlaProGluHisGluGly 1326
Db   5200 GGCACCAGCGTCCCTGGAGAACCAGCCTCACCTATCTCCCA-GCGCCTGCCCCCCAAGGT 5258

Qy   1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeu 1346
Db   5259 AGAATCCCTGGAGAGTCTCTACTTCACTCCCATCCCTGCTCGGAGTCAGGCCCCCCT--- 5315

Qy   1347 GlySerLeuGlyGlySerProSerLeuPro------------------------------ 1356
Db   5316 ---------GGAGAGCAGCCTGGACTCCCTGGGAGACGTCTTCCTGGACTCGGGTCGTAA 5366

Qy   1357 ------------------------------------------------AspGlnAspLys 1360
Db   5367 GACCCGCTCCGCTCGTCGGCGCACCACGCAGATCATCAACATCACCATGACCAAGAA--- 5423

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeu-AspThrThrValThrGlnThrAlaGl 1380
Db   5424 GCTAGATGTGGAAGAGCCAGACAGCGCCAACTCATCGTTCTACAGCACGCGGTCTGCTCC 5483

Qy   1380 uAlaValGluLysValIle-----------GluThrValValIleSerGluThrGlyGl 1396
Db   5484 TGCTTCCCAGGCTAGCCTGCGAGCCACCTCCTCTACTCAGTCTCTAGCTCGCCTGGGT-- 5541

Qy   1396 uSerProGluCysValGlyAlaHisLeuLeu----------------------------- 1406
Db   5542 -TCTCCCGATTATGGCAACTCAGCCCTGCTCAGCTTGCCTGGCTACCGCCCCACCACTCG 5600

Qy   1407 ----ProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAs 1425
Db   5601 CAGTTCTGCTCGTCGTTCCCAGGCCGG-GGTGTCCAGTGGGGCCCCTCCAGGAAGGAACA 5659
```

```
Qy     1425 pThrValPro----------------------------------LeuGlyProGluSe 1433
                 :::                                      |||   |||
Db     5660 GCTTCTACATGGGCACTTGCCAGGATGAGCCTGAGCAGCTGGATGACTGGAACCGCATTG 5719

Qy     1433 rGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisProAs 1453
              :::|||   :::            |||   |||      |||     |||   |||
Db     5720 CAGAGCTGCAGCAGCGCAATCGAGTGTGCCCCCCACATCTGAAGACCTGCTATCCCCTGG 5779

Qy     1453 pLeuGlnGlyGluIleSerAlaSerGlnArgGluArg---SerGluGluGluAspLysPr 1472
              :::|||                              |||     :::|||     :::||
Db     5780 AGTCCAGGCCTTCCCTGAGCCTGGGCACCATCACAGATGAGGAGATGAAAACTGGAGACC 5839

Qy     1472 oAsp-------------------------AlaGlyProAspAlaAspGly 1480
             |                            |||   |||||:::   |||
Db     5840 CCCAAGAGACCCTGCGCCGAGCCAGCATGCAGCCAATCCAGATAGCCGAGGG 5891
```

RESULT 9

```
US-08-467-781-3
; Sequence 3, Application US/08467781
; Patent No. 5780596
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/467,781
;      FILING DATE:  06-JUN-1995
;      CLASSIFICATION:  424
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6306 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..6306
;    PUBLICATION INFORMATION:
;      AUTHORS:  COMPTON, DUANE A
;      AUTHORS:  SZILAK, ILYA
;      AUTHORS:  CLEVELAND, DON W
;      TITLE:  PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
;      TITLE:  PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
;      TITLE:  SEGREGATION OF PROTEINS AT MITOSIS
;      JOURNAL:  J. Cell Biol.
;      VOLUME:  116
;      PAGES:  1395-1408
;      DATE:  1992
US-08-467-781-3
```

Alignment Scores:

```
Pred. No.:              1.98e-19      Length:          6306
Score:                  415.00        Matches:         373
Percent Similarity:     35.28%        Conservative:    255
Best Local Similarity:  20.96%        Mismatches:      663
Query Match:            5.14%         Indels:          493
DB:                     1             Gaps:            72

US-09-902-432-4 (1-1596) x US-08-467-781-3 (1-6306)

Qy       28 GlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspPro 47
             ||||||          |||      :::|||   |||||:::   |||         |||
Db     1050 GGCCACTCAGGAGTGGCTAGAGAAGCAGGCCCAGCTGGAGAAGGAGCTCAGCGCAGCCCT 1109

Qy       48 Ala-ThrLys------LeuProGlnLysAsn------------GlyGlnLeuSerSerVa 61
             |||   |||       |||   :::||||||            |||:::||||||   ::
Db     1110 GCAGGACAAGAAATGCCTTGAAGAGAAGAACGAAATCCTTCAGGGAAAACTTTCACAGCT 1169

Qy       61 l---------------------------AsnGlyValAlaGluGlnGlyAspVal----- 70
             :                           |||      |||:::|||:::|||
Db     1170 GGAAGAACACTTGTCCCAGCTGCAGGATAACCCACCCCAGGAGAAGGGCGAGGTGCTGGG 1229

Qy       71 ----------HisValGlnGluGluAsnGlnGluGly---------------------- 79
                       ::::::           ||||||
Db     1230 TGATGTCTTGCAGCTGGAAACCTTGAAGCAAGAGGCAGCCACTCTTGCTGCAAACAACAC 1289

Qy       80 ----------GlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAs 96
                                |||:::::::::   :::   ||||||:::|||::::::
Db     1290 ACAGCTCCAAGCCAGGGTAGAGATGCTGGAGACTGAGCGAGGCCAGCAGGAAGCCAAGCT 1349

Qy       96 pValArgGluLysAspArgValGluGlu------------------------------ 105
              :::   |||:::         ||||||
Db     1350 GCTTGCTGAGCGGGGCCACTTCGAAGAAGAAAAGCAGCAGCTGTCTAGCCTGATCACTGA 1409

Qy      106 ------------MetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGl 121
                        ::::::      |||   |||   |||::::::   :::      ||
Db     1410 CCTGCAGAGCTCCATCTCCAACCTCAGCCAGGCCAAGGAAGAGCTGGAGCAGGCCTCCCA 1469

Qy      121 nGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMe 141
             |          :::   :::      |||:::   :::      :::   :::      ::
Db     1470 GGCTCATGGGGCCCGGTTGACTGCCCAGGTGGCCTCTCTGACCTCTGAGCTCACCACACT 1529

Qy      141 tValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheVa 161
             :           :::   |||      ::::::
Db     1530 CAATGCCACCATCCAGCAACAGGATCAAGAACTG-----------------------GC 1565

Qy      161 lGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLe 181
             |||   |||      |||:::      :::      :::   :::   |||:::|||
Db     1566 TGGCCTGAAGCAGCAGGCCAAAGAGAAGCAGGCCCAGCTAGCACAGACCCTCCAA----- 1620

Qy      181 uThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGl 201
                                    :::      :::||||||   |||         |||||
Db     1621 ---------------------CAGCAAGAACAGGCCTCCCAGGGCCTC---CGCCACCA 1655

Qy      201 nGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGl 221
             |         ||||||                     ||||||   :::||||||::
Db     1656 G---------GTGGAGCAGCTAAGCAGTAGCCTGAAGCAGAAGGAGCAGCAGTTGAAGGA 1706

Qy      221 nSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLe 241
             :      ||||||||||||   |||         ::::::      :::   ::::::
Db     1707 GGTAGCGGAGAAGCAGGAGGCAACT------AGGCAGGACCATGCCCAGCAACTGGCCAC 1760

Qy      241 uGlnAlaGluSerAspGlnAlaAla------GluGluGluAlaLysAspGluGlyGluGl 259
                ||||||         :::|||:::      |||   :::|||      :::      ||
Db     1761 TGCTGCAGAGGAGCGAGAGGCCTCCTTAAGGGAGCGGGATGCGGCTCTCAAGCAGCTGGA 1820

Qy      259 uLysGlnGluLysGluProThrLysSerProGlu------SerProSerSerProValAs 277
             |      |||||||||         |||                         |||.
Db     1821 GGCACTGGAGAAGGAGAAGGCTGCCAAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGC 1880

Qy      277 nSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLy 297
              :::|||      |||   :::            |||                  |||::
Db     1881 TAATGAAGCCCGGGACAGTGCCCAGACCTCAGTGACACAG---------GCCCAGCGGGA 1931
```

```
Qy     297 sLysThrSerPheLysLysSerLysGluAsp---------AspLeuGluThrAlaGluLy 314
           :||           :::       |||:::          :::|||||||||    ::
Db    1932 GAAGGCAGAGCTGAGCCGGAAGGTGGAGGAACTCCAGGCCTGTGTTGAGACAGCCCGCCA 1991

Qy     314 sArgLysGluGlnGluAlaGluLysValAspGluGluGlu---------LysGluLysTh 331
           :   :::    :::          :::|||   |||   |||         :::    :::
Db    1992 GGAACAGCATGAGGCCCAGGCCCAGGTTGCAGAGCTAGAGTTGCAGCTGCGGTCTGAGCA 2051

Qy     331 rGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAl 351
           :::      |||:::|||:::::          |||::::::     ||||||     |||
Db    2052 GCAAAAAGCAACTGAGAAAGAAAGGGTGGCCAGGAGAAGGACCAG------CTCCAGGA 2105

Qy     351 aAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSe 371
             :::    ::::::          |||:::            |||   ||||||
Db    2106 GCAGCTCCAGGCCCTCAAAGAGTCCTTGAAGGTCACCAAGGGCAGCCTTGAA-------- 2157

Qy     371 rGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGl 391
            ||||||||               |||    :::       :::|||:::
Db    2158 -GAGGAGAAG------CGCAGGGCTGCAGATGCCCTGGAAGAGCAGCAGCGTTGTATCTC 2210

Qy     391 nGluValValAlaGluVal-----------------HisValSerThrValGluLysTh 405
            |||:::    ||||||                   |||           ::::::
Db    2211 TGAGCTGAAGGCAGAGACCCGAAGCCTGGTGGAGCAGCATAAGCGGGAACGAAAGGAGCT 2270

Qy     405 rGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGl 425
            |||||||||:::    |||              |||         ::::::       |||
Db    2271 GGAAGAAGAGAGGGCTGGGCGCAAGGGGCTGGAGGCTCGATTACTGCAGCTTGGGGAGGC 2330

Qy     425 yGluSerLeuProProGlu-----------LysLeuAlaGluProGlnGluValProGl 441
                       |||               :::|||||||||
Db    2331 CCATCAGGCTGAGACTGAAGTCCTGCGGCGGGAGCTGGCAGAGGCCATGGCTGCCCAGCA 2390

Qy     441 nGluAlaGlu---ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGl 460
              ||||||       |||:::|||:::|||                     |||
Db    2391 CACAGCTGAGAGTGAGTGTGAGCAGCTCGTCAAAGAAGTAGCTGCCTGGCGTGACGGGTA 2450

Qy     460 yAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGl 480
           :::     :::|||                   ||||||   |||:::      :::|||
Db    2451 TGAGGATAGCCAG--------------CAAGAGGAGGC-ACAGTATGGCGCCATGTTCC 2494

Qy     480 uGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySe 500
            |||          :::    |||
Db    2495 AGGAACAGCTGATGACTTTGAA-------------------------------------- 2516

Qy     500 rProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLy 520
                                                             |||      ::
Db    2517 ----------------------------------------------GGAGGAATGTGAGA 2530

Qy     520 sGlyLysArgGlyGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGl 540
           :||      |||      |||    :::        |||         |||
Db    2531 AGGCCCGCCAGGAGCTGCAGGAGGCAAAGGAGAAGGTGGCAGGCATAGAATCCCACAGCG 2590

Qy     540 uSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPr 560
           :::|||         |||:::           |||   ::::::    :::|||    :::||
Db    2591 AGCTCCA---------GATAAGCCGGCAGCAGAACAAACTAGCTGAGCTCCATGCCAACC 2641

Qy     560 oGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAla-- 579
           |   ::::::|||           ||||||    |||       |||      ||||||
Db    2642 TGGCCAGAGCACT------CCAGCAGGTCCAAGAGAAGGAAGTCAGGGCCCAGAAGCTTG 2695

Qy     580 -----------------------GluGluGly-ThrThrSerAspGlyGluLysLysA 591
                                  ::::::|||      ||||||           :::
Db    2696 CAGATGACCTCTCCACTCTGCAGGAAAAGATGGCTGCCACCAGCAAAGAGGTGGCCCGCT 2755

Qy     591 rgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValA 611
             |||    :::      |||          ::::::              ::::::   |||:
Db    2756 TGGAGACCTTGGTCCGCAAGGCAGGTGAGCAGCAGGAAACAGCCTCCCGGGAGTTAGTCA 2815

Qy     611 rgArgProSerGluSer---AspLysGluGluGlu---LeuGluLysValLysSerAlaT 629
            ::    |||:::    :::     |||::::::     |||    ||||||:::     :::
Db    2816 AGGAGCCTGCGAGGGCAGGAGACAGACAGCCCGAGTGGCTGGAAGAGCAACAGGGACGCC 2875

Qy     629 hrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyG 649
                 |||||    :::    :::        |||:::    |||     :::     :::|||
Db    2876 AGTTCTGCAGCACACAGGCAGCGCTGCAGGCTATGGAGCGGGAGGCAGAGCAGATGGGCA 2935

Qy     649 luGlu--------------------------------------GlnLysProGluGlu- 655
             |||                                       |||:::    ||||||
Db    2936 ATGAGCTGGAACGGCTGCGGGCCGCGCTGATGGAGAGCCAGGGGCAGCAGCAGGAGGAGC 2995

Qy     655 ------------------------------------------------------------ 655

Db    2996 GTGGGCAGCAGGAAAGGGAGGTGGCGCGGCTGACCCAGGAGCGGGGCCGTGCCCAGGCTG 3055

Qy     656 -----------------------------------------------------------P 656

Db    3056 ACCTTGCCCTGGAGAAGGCGGCCAGAGCAGAGCTTGAGATGCGGCTGCAGAACGCCCTCA 3115

Qy     656 roLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerL 676
              ::::::||||||:::    :::             |||||||||          :::    |
Db    3116 ACGAGCAGCGTGTGGAGTTCGCTACCCTGCAAGAGGCACTGGCTCATGCCCTGACGGAAA 3175

Qy     676 ysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyG 696
             |||:::                              |||   |||   |||      :::    |
Db    3176 AGGAA-----------------GGCAAGGACCAGGAGTT-GGCCAAGCTTCGTGGTCTG 3216

Qy     696 lyAspSer------HisArgAlaGluGluAlaSerLysAspLysGluAla------GlyT 712
             ||  |||         ||||||          |||    ::::::||||||       ||||
Db    3217 GAGGCAGCCCAGATAAAAGAGCTGGAGGAACTTCGGCAAACCGTGAAGCAACTGAAGGAA 3276

Qy     712 hrAspAlaValProAla-SerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPro 731
             ||               ||| :::    :::|||::::::      |||      ||||||     :::
Db    3277 CA-----------GCTGGCTAAGAAAGAAAAGGAGCACGCATCTGGCTCAGGAGCCCAA 3324

Qy     732 GluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArg 751
                 |||          :::|||        |||         |||:::     :::
Db    3325 TCTGAGGCTGCTGGCAGGACAGAGCCAACAGGCCCCAAGCTG---GAAGCACTGCGGGCA 3381

Qy     752 LeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGlu--------- 768
              |||:::         ::::::            |||    :::|||:::|||:::
Db    3382 GAGGTGAGCAAGCTGGAACAGCAATGCCAGAAGCAGCAGGAGCAGGCTGACAGCCTGGAA 3441

Qy     769 --------------------------------------AspSerSerValGluGlnLeu 775
                                                 ||||||::::::|||    |||
Db    3442 CGCAGCCTCGAGGCTGAGCGGGCCTCCCGGGCTGAGCGGGACAGTGCTCTGGAGACTCTG 3501

Qy     776 SerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIlePro 795
              ::::::|||          :::|||         |||
Db    3502 CAGGGCCAGTTAGAGGAGAAGGCCCAGGAGCTAGGGCACAGTCAGAGTGCCTTAGCCTCG 3561

Qy     796 GlyArgArgLysLysArgAla------AspGlyLysGlnGluGlnAlaThrValGluAsp 813
             :::|||:::          |||          :::          ::::::||||||
Db    3562 GCCCAACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAAGACCACAGCAAGGCTGAAGAT 3621

Qy     814 SerGlyProValGluIle------------AsnGluAspAspProAsnValProAlaVal 829
                       ::::::             |||    :::              :::
Db    3622 GAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCTGAGAGGAAAAATAGCCTCATC 3681

Qy     830 ValProLeu---SerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsn 848
                 |||      |||    :::    :::    |||:::    :::|||   :::|||
Db    3682 AGCAGCTTGGAGGAGGAGGTGTCCATCCTGAATCGCCAGGTCCTGGAGAAGGAGGGG--- 3738

Qy     849 ThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeu 868
              |||          :::|||          |||                |||    |||:::    |||
Db    3739 ---GAGAGCAAGGAGTTGAAGCGGCTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTG 3795

Qy     869 ValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGlu 888
                       :::    :::                                  :::
Db    3796 GAGGAGAGCTGCGCCTGCTGC---------------------------------AGGCAG 3822

Qy     889 ArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGlu 908
            |||   |||:::
Db    3823 AGACAGCCAGCA------------------------------------------------ 3834

Qy     909 AlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVal 928
                       |||    |||:::          ::::::
```

```
Db   3835 ---------ACAGTGCCAGAGCTGCAGAACGCAGCTCTGCTCTGCGGGAGG--------- 3876

Qy    929 LeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGlu--- 947
Db   3877 ------AGGTGCAGAGCCTCCGGGAGGGAGGCTGAGAAACAGCGGGTGGCTTCAGAGAAC 3930

Qy    948 ------ValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArg 965
Db   3931 CTGCGGCAGGAGCTGACCTCACAGGCTGAGCGTGCGGAGGAGCTGGGCCAAGAATTGAAG 3990

Qy    966 Thr------GluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSer 983
Db   3991 GCGTGGCAGGAGAAGTTCTTCCAGAAAGAGCAGGCCCTCTCCACCCTGCAGCTC------ 4044

Qy    984 AlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGlu 1003
Db   4045 ---------------------------------------GAGCACACCAGCACACAGGCC 4065

Qy   1004 ValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGln 1023
Db   4066 CTGGTGAGTGAGCTGCTGCCAGCTAAGCACCTCTGCCAGCAGCTGCAGGCCGAGCAGGCC 4125

Qy   1024 AlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArg 1043
Db   4126 GCTGCCGAGAAACGCCACCGTGAGGAGCTGGAG------------CAGAGCAAGCAGGCC 4173

Qy   1044 ThrGlySerLysAlaLeuGlu--------------------------------------- 1050
Db   4174 GCTGGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAGCGGGAGCTTGGGGAGCTGATTCCT 4233

Qy   1051 ---------LysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys--- 1066
Db   4234 CTGCGGCAGAAGGTGGCAGAGCAGGAGCGAACAGCTCAGCAGCTGCGGGCAGAGAAGGCC 4293

Qy   1066 ------------------------------------------------------------ 1066
Db   4294 AGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTGCTGGCAGAGGAGAAC 4353

Qy   1066 ------------------------------------------------------------ 1066
Db   4354 CGGGGGCTGGGTGAGCGGGCCAACCTTGGCCGGCAGTTTCTGGAAGTGGAGTTGGACCAG 4413

Qy   1067 ------GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeu 1084
Db   4414 GCCCGGGAAAAGTATGTCCAAGAGTTGGCAGCCGTACGTGCTGATGCTGAGACCCGTCTG 4473

Qy   1085 AlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThr 1104
Db   4474 GCTGAGGTGCAGCGAGAAGCACAGAGCACTGCCCGGGAGCTGGAGGTG------ATGACT 4527

Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGln 1124
Db   4528 GCCAAGTATGAGGGTGCC------------------AAGGTCAAGGTCCTGGAGGAGAGG 4569

Qy   1125 AlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrPro 1144
Db   4570 CAGCGGTTCCAGGAAGAGAGGCAGAAACTCACTGCCCAGGTGGAAGAACTGAGTAAGAAA 4629

Qy   1145 LeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAsp------GluThrIleAspSer 1162
Db   4630 CTGGCTGACTCTGACCAAGCCAGCAAGGTGCAGCAGCAGAAGCTGAAGGCTGTCCAGGCT 4689

Qy   1163 Gln------AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
Db   4690 CAGGGAGGCGAGAGCCAGCAGGAGGCCCAGCGCTTCCAGGCCCAGCTGAATGAACTGCAA 4749

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSer------------------------ThrLeu 1192
Db   4750 GCCCAGTTGAGCCAGAAGGAGCAGGCAGCTGAGCACTATAAGCTGCAGATGGAGAAAGCC 4809

Qy   1193 ProAsnAsnValProAlaGlnGluGluHisGlyGluGlu---------------ProGly 1207
Db   4810 AAAACACATTATGATGCCAAGAAGCAGCAGAACCAAGAGCTGCAGGAGCAGCTGCGGAGC 4869

Qy   1208 ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAla 1227
Db   4870 CTGGAGCAGCTGCAGAAGGAAAACAAAGAGCTGCGAGCTGAAGCT--------------- 4914

Qy   1228 LysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu 1247
Db   4915 ---GAACGGCTGGGCCATGAGCTACAG---------CAGGCTGGGCTGAAGACCAAGGAG 4962

Qy   1248 GluGlnGluValPheValHis------SerGlyProAsnSerGlnLysAlaAlaAspVal 1265
Db   4963 GCTGAACAGACCTGCCGCCACCTTACTGCCCAGGTGCGCAGCCTGGAGGCACAGGTTGCC 5022

Qy   1266 ThrTyrAspSerGluValMetGlyValAlaGlyCysGln------------GluLysGlu 1281
Db   5023 CATGCAGACCAGCAGCTTCGAGACCTGGGCAAATTCCAGGTGGCAACTGATGCTTTAAAG 5082

Qy   1282 SerThrGluValGlnSer------------------------------------LeuSer 1289
Db   5083 AGCCGTGAGCCCCAGGCTAAGCCCCAGCTGGACTTGAGTATTGACAGCCTGGATCTGAGC 5142

Qy   1290 LeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGlu 1309
Db   5143 TGCGAGGAGGGGACCCCACTCAGTATCACCAGCAAGCTGCCTCGT---ACCCAGCCAGAC 5199

Qy   1310 GlnValSerGluGluGlyGluGlnGlu---------ThrAlaAlaProGluHisGluGly 1326
Db   5200 GGCACCAGCGTCCCTGGAGAACCAGCCTCACCTATCTCCCA-GCGCCTGCCCCCCAAGGT 5258

Qy   1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeu 1346
Db   5259 AGAATCCCTGGAGAGTCTCTACTTCACTCCCATCCCTGCTCGGAGTCAGGCCCCCCT--- 5315

Qy   1347 GlySerLeuGlyGlySerProSerLeuPro------------------------------ 1356
Db   5316 ---------GGAGAGCAGCCTGGACTCCCTGGGAGACGTCTTCCTGGACTCGGGTCGTAA 5366

Qy   1357 ------------------------------------------------AspGlnAspLys 1360
Db   5367 GACCCGCTCCGCTCGTCGGCGCACCACGCAGATCATCAACATCACCATGACCAAGAA--- 5423

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeu-AspThrThrValThrGlnThrAlaGl 1380
Db   5424 GCTAGATGTGGAAGAGCCAGACAGCGCCAACTCATCGTTCTACAGCACGCGGTCTGCTCC 5483

Qy   1380 uAlaValGluLysValIle------------GluThrValValIleSerGluThrGlyGl 1396
Db   5484 TGCTTCCCAGGCTAGCCTGCGAGCCACCTCCTCTACTCAGTCTCTAGCTCGCCTGGGT-- 5541

Qy   1396 uSerProGluCysValGlyAlaHisLeuLeu----------------------------- 1406
Db   5542 -TCTCCCGATTATGGCAACTCAGCCCTGCTCAGCTTGCCTGGCTACCGCCCCACCACTCG 5600

Qy   1407 ----ProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAs 1425
Db   5601 CAGTTCTGCTCGTCGTTCCCAGGCCGG-GGTGTCCAGTGGGGCCCCTCCAGGAAGGAACA 5659

Qy   1425 pThrValPro------------------------------------LeuGlyProGluSe 1433
Db   5660 GCTTCTACATGGGCACTTGCCAGGATGAGCCTGAGCAGCTGGATGACTGGAACCGCATTG 5719

Qy   1433 rGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisProAs 1453
Db   5720 CAGAGCTGCAGCAGCGCAATCGAGTGTGCCCCCACATCTGAAGACCTGCTATCCCCTGG 5779

Qy   1453 pLeuGlnGlyGluIleSerAlaSerGlnArgGluArg---SerGluGluGluAspLysPr 1472
Db   5780 AGTCCAGGCCTTCCCTGAGCCTGGGCACCATCACAGATGAGGAGATGAAAACTGGAGACC 5839

Qy   1472 oAsp--------------------------AlaGlyProAspAlaAspGly 1480
Db   5840 CCCAAGAGACCCTGCGCCGAGCCAGCATGCAGCCAATCCAGATAGCCGAGGG 5891
```

```
RESULT 10
US-08-483-924-3
; Sequence 3, Application US/08483924
; Patent No. 5882876
;  GENERAL INFORMATION:
;    APPLICANT:  TOUKATLY, GARY
;    APPLICANT:  LIDGARD, GRAHAM P
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA, HURWITZ & THIBEAULT
;      STREET:  125 HIGH STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02110
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/483,924
;      FILING DATE:  07-JUN-1995
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (617) 248-7000
;      TELEFAX:  (617) 248-7100
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6306 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..6306
;    PUBLICATION INFORMATION:
;      AUTHORS:  COMPTON, DUANE A
;      AUTHORS:  SZILAK, ILYA
;      AUTHORS:  CLEVELAND; DON W
;      TITLE:  PRIMARY STRUCTURE OF NUMA, AN INTRANUCLEAR
;      TITLE:  PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
;      TITLE:  SEGREGATION OF PROTEINS AT MITOSIS
;      JOURNAL:  J. Cell Biol.
;      VOLUME:  116
;      PAGES:  1395-1408
;      DATE:  1992
US-08-483-924-3

Alignment Scores:
Pred. No.:               1.98e-19      Length:         6306
Score:                   415.00        Matches:        373
Percent Similarity:      35.28%        Conservative:   255
Best Local Similarity:   20.96%        Mismatches:     663
Query Match:             5.14%         Indels:         493
DB:                      2             Gaps:           72

US-09-902-432-4 (1-1596) x US-08-483-924-3 (1-6306)

Qy      28 GlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspPro 47
             ||||||         |||      :::|||  ||||||:::    |||         |||
Db    1050 GGCCACTCAGGAGTGGCTAGAGAAGCAGGCCCAGCTGGAGAAGGAGCTCAGCGCAGCCCT 1109

Qy      48 Ala-ThrLys------LeuProGlnLysAsn-----------GlyGlnLeuSerSerVa 61
             |||   |||      |||    :::||||||           |||:::||||||    ::
Db    1110 GCAGGACAAGAAATGCCTTGAAGAGAAGAACGAAATCCTTCAGGGAAAACTTTCACAGCT 1169
```

```
Qy      61 l--------------------------AsnGlyValAlaGluGlnGlyAspVal----- 70
           :                          |||        |||:::|||:::|||
Db    1170 GGAAGAACACTTGTCCCAGCTGCAGGATAACCCACCCCAGGAGAAGGGCGAGGTGCTGGG 1229

Qy      71 ----------HisValGlnGluGluAsnGlnGluGly---------------------- 79
                     ::::::         ||||||
Db    1230 TGATGTCTTGCAGCTGGAAACCTTGAAGCAAGAGGCAGCCACTCTTGCTGCAAACAACAC 1289

Qy      80 ----------GlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAs 96
                               |||:::::::::   :::   ||||||:::|||::::::
Db    1290 ACAGCTCCAAGCCAGGGTAGAGATGCTGGAGACTGAGCGAGGCCAGCAGGAAGCCAAGCT 1349

Qy      96 pValArgGluLysAspArgValGluGlu-------------------------------- 105
           :::   |||:::         ||||||
Db    1350 GCTTGCTGAGCGGGGCCACTTCGAAGAAGAAAAGCAGCAGCTGTCTAGCCTGATCACTGA 1409

Qy     106 ------------MetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGl 121
                       ::::::      |||   |||   |||::::::   :::      ||
Db    1410 CCTGCAGAGCTCCATCTCCAACCTCAGCCAGGCCAAGGAAGAGCTGGAGCAGGCCTCCCA 1469

Qy     121 nGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMe 141
           |       :::   :::      |||:::      :::   :::   :::      ::
Db    1470 GGCTCATGGGGCCCGGTTGACTGCCCAGGTGGCCTCTCTGACCTCTGAGCTCACCACACT 1529

Qy     141 tValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheVa 161
           :               :::   |||      ::::::
Db    1530 CAATGCCACCATCCAGCAACAGGATCAAGAACTG-----------------------GC 1565

Qy     161 lGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLe 181
             |||   |||         |||:::   :::   :::   :::   |||:::|||
Db    1566 TGGCCTGAAGCAGCAGGCCAAAGAGAAGCAGGCCCAGCTAGCACAGACCCTCCAA----- 1620

Qy     181 uThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGl 201
                                 :::      :::||||||   |||         |||||
Db    1621 ---------------------CAGCAAGAACAGGCCTCCCAGGGCCTC---CGCCACCA 1655

Qy     201 nGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGl 221
           |         ||||||                     ||||||   :::||||||::
Db    1656 G---------GTGGAGCAGCTAAGCAGTAGCCTGAAGCAGAAGGAGCAGCAGTTGAAGGA 1706

Qy     221 nSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLe 241
           :      ||||||||||||   |||      ::::::      :::   ::::::
Db    1707 GGTAGCGGAGAAGCAGGAGGCAACT------AGGCAGGACCATGCCCAGCAACTGGCCAC 1760

Qy     241 uGlnAlaGluSerAspGlnAlaAla------GluGluGluAlaLysAspGluGlyGluGl 259
              ||||||      :::|||:::      |||   :::|||      :::      ||
Db    1761 TGCTGCAGAGGAGCGAGAGGCCTCCTTAAGGGAGCGGGATGCGGCTCTCAAGCAGCTGGA 1820

Qy     259 uLysGlnGluLysGluProThrLysSerProGlu------SerProSerSerProValAs 277
           |      |||||||||                  |||                |||
Db    1821 GGCACTGGAGAAGGAGAAGGCTGCCAAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGC 1880

Qy     277 nSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLy 297
           :::|||         |||   :::         |||                |||::
Db    1881 TAATGAAGCCCGGGACAGTGCCCAGACCTCAGTGACACAG---------GCCCAGCGGGA 1931

Qy     297 sLysThrSerPheLysLysSerLysGluAsp---------AspLeuGluThrAlaGluLy 314
           :|||         :::      |||:::             :::|||||||||   ::
Db    1932 GAAGGCAGAGCTGAGCCGGAAGGTGGAGGAACTCCAGGCCTGTGTTGAGACAGCCCGCCA 1991

Qy     314 sArgLysGluGlnGluAlaGluLysValAspGluGluGlu---------LysGluLysTh 331
           :   :::   :::      :::|||   |||   |||            :::   :::
Db    1992 GGAACAGCATGAGGCCCAGGCCCAGGTTGCAGAGCTAGAGTTGCAGCTGCGGTCTGAGCA 2051

Qy     331 rGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAl 351
              :::   |||:::|||::::::         |||::::::   ||||||      |||
Db    2052 GCAAAAAGCAACTGAGAAAGAAAGGGTGGCCCAGGAGAAGGACCAG------CTCCAGGA 2105

Qy     351 aAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSe 371
                 :::   ::::::         |||:::         |||   ||||||
Db    2106 GCAGCTCCAGGCCCTCAAAGAGTCCTTGAAGGTCACCAAGGGCAGCCTTGAA-------- 2157
```

```
Qy    371 rGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGl 391
Db   2158 -GAGGAGAAG------CGCAGGGCTGCAGATGCCCTGGAAGAGCAGCAGCGTTGTATCTC 2210

Qy    391 nGluValValAlaGluVal------------------HisValSerThrValGluLysTh 405
Db   2211 TGAGCTGAAGGCAGAGACCCGAAGCCTGGTGGAGCAGCATAAGCGGGAACGAAAGGAGCT 2270

Qy    405 rGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGl 425
Db   2271 GGAAGAAGAGAGGGCTGGGCGCAAGGGGCTGGAGGCTCGATTACTGCAGCTTGGGGAGGC 2330

Qy    425 yGluSerLeuProProGlu------------LysLeuAlaGluProGlnGluValProGl 441
Db   2331 CCATCAGGCTGAGACTGAAGTCCTGCGGCGGGAGCTGGCAGAGGCCATGGCTGCCCAGCA 2390

Qy    441 nGluAlaGlu---ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGl 460
Db   2391 CACAGCTGAGAGTGAGTGTGAGCAGCTCGTCAAAGAAGTAGCTGCCTGGCGTGACGGGTA 2450

Qy    460 yAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGl 480
Db   2451 TGAGGATAGCCAG---------------CAAGAGGAGGC-ACAGTATGGCGCCATGTTCC 2494

Qy    480 uGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySe 500
Db   2495 AGGAACAGCTGATGACTTTGAA-------------------------------------- 2516

Qy    500 rProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLy 520
Db   2517 ----------------------------------------------GGAGGAATGTGAGA 2530

Qy    520 sGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGl 540
Db   2531 AGGCCCGCCAGGAGCTGCAGGAGGCAAAGGAGAAGGTGGCAGGCATAGAATCCCACAGCG 2590

Qy    540 uSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPr 560
Db   2591 AGCTCCA---------GATAAGCCGGCAGCAGAACAAACTAGCTGAGCTCCATGCCAACC 2641

Qy    560 oGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAla-- 579
Db   2642 TGGCCAGAGCACT------CCAGCAGGTCCAAGAGAAGGAAGTCAGGGCCCAGAAGCTTG 2695

Qy    580 ------------------------GluGluGly-ThrThrSerAspGlyGluLysLysA 591
Db   2696 CAGATGACCTCTCCACTCTGCAGGAAAAGATGGCTGCCACCAGCAAAGAGGTGGCCCGCT 2755

Qy    591 rgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValA 611
Db   2756 TGGAGACCTTGGTCCGCAAGGCAGGTGAGCAGCAGGAAACAGCCTCCCGGGAGTTAGTCA 2815

Qy    611 rgArgProSerGluSer---AspLysGluGluGlu---LeuGluLysValLysSerAlaT 629
Db   2816 AGGAGCCTGCGAGGGCAGGAGACAGACAGCCCGAGTGGCTGGAAGAGCAACAGGACGCC 2875

Qy    629 hrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyG 649
Db   2876 AGTTCTGCAGCACACAGGCAGCGCTGCAGGCTATGGAGCGGGAGGCAGAGCAGATGGGCA 2935

Qy    649 luGlu--------------------------------------GlnLysProGluGlu- 655
Db   2936 ATGAGCTGGAACGGCTGCGGGCCGCGCTGATGGAGAGCCAGGGGCAGCAGCAGGAGGAGC 2995

Qy    655 ------------------------------------------------------------ 655
Db   2996 GTGGGCAGCAGGAAAGGGAGGTGGCGCGGCTGACCCAGGAGCGGGGCCGTGCCCAGGCTG 3055

Qy    656 -----------------------------------------------------------P 656
Db   3056 ACCTTGCCCTGGAGAAGGCGGCCAGAGCAGAGCTTGAGATGCGGCTGCAGAACGCCCTCA 3115

Qy    656 roLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerL 676
Db   3116 ACGAGCAGCGTGTGGAGTTCGCTACCCTGCAAGAGGCACTGGCTCATGCCCTGACGGAAA 3175

Qy    676 ysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyG 696
Db   3176 AGGAA-----------------GGCAAGGACCAGGAGTT-GGCCAAGCTTCGTGGTCTG 3216

Qy    696 lyAspSer------HisArgAlaGluGluAlaSerLysAspLysGluAla------GlyT 712
Db   3217 GAGGCAGCCCAGATAAAAGAGCTGGAGGAACTTCGGCAAACCGTGAAGCAACTGAAGGAA 3276

Qy    712 hrAspAlaValProAla-SerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPro 731
Db   3277 CA-----------GCTGGCTAAGAAAGAAAAGGAGCACGCATCTGGCTCAGGAGCCCAA 3324

Qy    732 GluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArg 751
Db   3325 TCTGAGGCTGCTGGCAGGACAGAGCCAACAGGCCCCAAGCTG---GAAGCACTGCGGGCA 3381

Qy    752 LeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGlu--------- 768
Db   3382 GAGGTGAGCAAGCTGGAACAGCAATGCCAGAAGCAGCAGGAGCAGGCTGACAGCCTGGAA 3441

Qy    769 -------------------------------------AspSerSerValGluGlnLeu 775
Db   3442 CGCAGCCTCGAGGCTGAGCGGGCCTCCCGGGCTGAGCGGGACAGTGCTCTGGAGACTCTG 3501

Qy    776 SerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIlePro 795
Db   3502 CAGGGCCAGTTAGAGGAGAAGGCCCAGGAGCTAGGGCACAGTCAGAGTGCCTTAGCCTCG 3561

Qy    796 GlyArgArgLysLysArgAla------AspGlyLysGlnGluGlnAlaThrValGluAsp 813
Db   3562 GCCCAACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAAGACCACAGCAAGGCTGAAGAT 3621

Qy    814 SerGlyProValGluIle------------AsnGluAspAspProAsnValProAlaVal 829
Db   3622 GAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCTGAGAGGAAAAATAGCCTCATC 3681

Qy    830 ValProLeu---SerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsn 848
Db   3682 AGCAGCTTGGAGGAGGAGGTGTCCATCCTGAATCGCCAGGTCCTGGAGAAGGAGGGG--- 3738

Qy    849 ThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeu 868
Db   3739 ---GAGAGCAAGGAGTTGAAGCGGCTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTG 3795

Qy    869 ValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGlu 888
Db   3796 GAGGAGAGCTGCGCCTGCTGC-------------------------------AGGCAG 3822

Qy    889 ArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGlu 908
Db   3823 AGACAGCCAGCA------------------------------------------------ 3834

Qy    909 AlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVal 928
Db   3835 ---------ACAGTGCCAGAGCTGCAGAACGCAGCTCTGCTCTGCGGGAGG--------- 3876

Qy    929 LeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGlu--- 947
Db   3877 ------AGGTGCAGAGCCTCCGGGAGGGAGGCTGAGAAACAGCGGGTGGCTTCAGAGAAC 3930

Qy    948 ------ValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArg 965
Db   3931 CTGCGGCAGGAGCTGACCTCACAGGCTGAGCGTGCGGAGGAGCTGGGCCAAGAATTGAAG 3990

Qy    966 Thr------GluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSer 983
Db   3991 GCGTGGCAGGAGAAGTTCTTCCAGAAAGAGCAGGCCCTCTCCACCCTGCAGCTC------ 4044

Qy    984 AlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGlu 1003
```

```
Db   4045 ----------------------------------------GAGCACACCAGCACACAGGCC 4065

Qy   1004 ValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGln 1023
          :::   |||   :::|||       :::          :::|||     ||||||
Db   4066 CTGGTGAGTGAGCTGCTGCCAGCTAAGCACCTCTGCCAGCAGCTGCAGGCCGAGCAGGCC 4125

Qy   1024 AlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArg 1043
          |||         :::   :::||||||   :::                |||:::   |||
Db   4126 GCTGCCGAGAAACGCCACCGTGAGGAGCTGGAG------------CAGAGCAAGCAGGCC 4173

Qy   1044 ThrGlySerLysAlaLeuGlu--------------------------------------- 1050
             |||         |||
Db   4174 GCTGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAGCGGGAGCTTGGGGAGCTGATTCCT 4233

Qy   1051 ---------LysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys--- 1066
                   ||||||   |||   |||      ::::::   |||   :::||||||
Db   4234 CTGCGGCAGAAGGTGGCAGAGCAGGAGCGAACAGCTCAGCAGCTGCGGGCAGAGAAGGCC 4293

Qy   1066 ------------------------------------------------------------ 1066

Db   4294 AGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTGCTGGCAGAGGAGAAC 4353

Qy   1066 ------------------------------------------------------------ 1066

Db   4354 CGGGGGCTGGGTGAGCGGGCCAACCTTGGCCGGCAGTTTCTGGAAGTGGAGTTGGACCAG 4413

Qy   1067 ------GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeu 1084
                ||||||   |||               |||:::               |||
Db   4414 GCCCGGGAAAAGTATGTCCAAGAGTTGGCAGCCGTACGTGCTGATGCTGAGACCCGTCTG 4473

Qy   1085 AlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThr 1104
          |||:::                  |||:::|||         |||||||||      :::|||
Db   4474 GCTGAGGTGCAGCGAGAAGCACAGAGCACTGCCCGGGAGCTGGAGGTG------ATGACT 4527

Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGln 1124
          |||      :::                  |||::::::   |||   |||:::
Db   4528 GCCAAGTATGAGGGTGCC------------------AAGGTCAAGGTCCTGGAGGAGAGG 4569

Qy   1125 AlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrPro 1144
                      |||      :::   ||||||                  |||
Db   4570 CAGCGGTTCCAGGAAGAGAGGCAGAAACTCACTGCCCAGGTGGAAGAACTGAGTAAGAAA 4629

Qy   1145 LeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAsp------GluThrIleAspSer 1162
          |||||||||||||||   |||         ||||||      :::   :::   :::
Db   4630 CTGGCTGACTCTGACCAAGCCAGCAAGGTGCAGCAGCAGAAGCTGAAGGCTGTCCAGGCT 4689

Qy   1163 Gln------AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          |||      :::|||:::         |||         |||:::|||:::   |||   :::
Db   4690 CAGGGAGGCGAGAGCCAGCAGGAGGCCCAGCGCTTCCAGGCCCAGCTGAATGAACTGCAA 4749

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSer----------------------ThrLeu 1192
          |||      :::||||||||||:::   :::
Db   4750 GCCCAGTTGAGCCAGAAGGAGCAGGCAGCTGAGCACTATAAGCTGCAGATGGAGAAAGCC 4809

Qy   1193 ProAsnAsnValProAlaGlnGluGluHisGlyGluGlu---------------ProGly 1207
                :::         |||:::::::::      :::|||
Db   4810 AAAACACATTATGATGCCAAGAAGCAGCAGAACCAAGAGCTGCAGGAGCAGCTGCGGAGC 4869

Qy   1208 ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAla 1227
             :::   |||:::         :::||||||   |||   |||
Db   4870 CTGGAGCAGCTGCAGAAGGAAAACAAAGAGCTGCGAGCTGAAGCT--------------- 4914

Qy   1228 LysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu 1247
                :::|||   |||   :::               |||   |||   ||||||
Db   4915 ---GAACGGCTGGGCCATGAGCTACAG---------CAGGCTGGGCTGAAGACCAAGGAG 4962

Qy   1248 GluGlnGluValPheValHis------SerGlyProAsnSerGlnLysAlaAlaAspVal 1265
          ::::::      |||         :::            :::|||
Db   4963 GCTGAACAGACCTGCCGCCACCTTACTGCCCAGGTGCGCAGCCTGGAGGCACAGGTTGCC 5022

Qy   1266 ThrTyrAspSerGluValMetGlyValAlaGlyCysGln-----------GluLysGlu 1281
             |||   ::::::   :::      |||                           :::
Db   5023 CATGCAGACCAGCAGCTTCGAGACCTGGGCAAATTCCAGGTGGCAACTGATGCTTTAAAG 5082

Qy   1282 SerThrGluValGlnSer------------------------------------LeuSer 1289
          |||   |||   |||:::                                    ||||||
Db   5083 AGCCGTGAGCCCCAGGCTAAGCCCCAGCTGGACTTGAGTATTGACAGCCTGGATCTGAGC 5142

Qy   1290 LeuGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGlu 1309
            |||||||||            :::         |||         |||   |||:::|||:::
Db   5143 TGCGAGGAGGGGACCCACTCAGTATCACCAGCAAGCTGCCTCGT---ACCCAGCCAGAC 5199

Qy   1310 GlnValSerGluGluGlyGluGlnGlu---------ThrAlaAlaProGluHisGluGly 1326
             |||      ||||||               :::   ||||||      :::|||
Db   5200 GGCACCAGCGTCCCTGGAGAACCAGCCTCACCTATCTCCCA-GCGCCTGCCCCCCAAGGT 5258

Qy   1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeu 1346
                         |||                  |||   ||||||   |||
Db   5259 AGAATCCCTGGAGAGTCTCTACTTCACTCCCATCCCTGCTCGGAGTCAGGCCCCCCT--- 5315

Qy   1347 GlySerLeuGlyGlySerProSerLeuPro------------------------------ 1356
                |||      |||   ||||||
Db   5316 ---------GGAGAGCAGCCTGGACTCCCTGGGAGACGTCTTCCTGGACTCGGGTCGTAA 5366

Qy   1357 ------------------------------------------------AspGlnAspLys 1360
                                                            ||||||:::
Db   5367 GACCCGCTCCGCTCGTCGGCGCACCACGCAGATCATCAACATCACCATGACCAAGAA--- 5423

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeu-AspThrThrValThrGlnThrAlaGl 1380
          |||   |||         :::   :::   |||               |||::::::|||
Db   5424 GCTAGATGTGGAAGAGCCAGACAGCGCCAACTCATCGTTCTACAGCACGCGGTCTGCTCC 5483

Qy   1380 uAlaValGluLysValIle------------GluThrValValIleSerGluThrGlyGl 1396
           |||   :::   :::                |||      :::::         |||
Db   5484 TGCTTCCCAGGCTAGCCTGCGAGCCACCTCCTCTACTCAGTCTCTAGCTCGCCTGGGT-- 5541

Qy   1396 uSerProGluCysValGlyAlaHisLeuLeu---------------------------- 1406
             ||||||:::         :::   ||||||
Db   5542 -TCTCCCGATTATGGCAACTCAGCCCTGCTCAGCTTGCCTGGCTACCGCCCCACCACTCG 5600

Qy   1407 ----ProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAs 1425
                |||   :::|||   |||   |||      |||
Db   5601 CAGTTCTGCTCGTCGTTCCCAGGCCGG-GGTGTCCAGTGGGGCCCCTCCAGGAAGGAACA 5659

Qy   1425 pThrValPro------------------------------------LeuGlyProGluSe 1433
             :::                                          |||   |||
Db   5660 GCTTCTACATGGGCACTTGCCAGGATGAGCCTGAGCAGCTGGATGACTGGAACCGCATTG 5719

Qy   1433 rGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisProAs 1453
           :::|||   :::               |||   |||      |||      |||   |||
Db   5720 CAGAGCTGCAGCAGCGCAATCGAGTGTGCCCCCCACATCTGAAGACCTGCTATCCCCTGG 5779

Qy   1453 pLeuGlnGlyGluIleSerAlaSerGlnArgGluArg---SerGluGluGluAspLysPr 1472
           :::|||                        |||         :::|||      :::||
Db   5780 AGTCCAGGCCTTCCCTGAGCCTGGGCACCATCACAGATGAGGAGATGAAAACTGGAGACC 5839

Qy   1472 oAsp--------------------------AlaGlyProAspAlaAspGly 1480
           |                            |||   ||||||:::   |||
Db   5840 CCCAAGAGACCCTGCGCCGAGCCAGCATGCAGCCAATCCAGATAGCCGAGGG 5891
```

RESULT 11
US-08-195-487-3
; Sequence 3, Application US/08195487
; Patent No. 5783403
; GENERAL INFORMATION:
; APPLICANT: TOUKATLY, GARY
; APPLICANT: LIDGARD, GRAHAM P
; TITLE OF INVENTION: NOVEL MALIGNANT CELL TYPE MARKERS OF THE
; TITLE OF INVENTION: INTERIOR NUCLEAR MATRIX
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TESTA HURWITZ & THIBEAULT
; STREET: 53 STATE STREET
; CITY: BOSTON
; STATE: MA
; COUNTRY: USA

```
;      ZIP:  02109
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.25
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/195,487
;       FILING DATE:
;       CLASSIFICATION:  435
;     PRIOR APPLICATION DATA:
;       APPLICATION NUMBER:  US/07/901,701
;       FILING DATE:
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  PITCHER ESQ, EDMUND R
;       REGISTRATION NUMBER:  27,829
;       REFERENCE/DOCKET NUMBER:  MTP-013
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  617/248-7000
;       TELEFAX:  617/248-7100
;   INFORMATION FOR SEQ ID NO:  3:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  6306 base pairs
;       TYPE:  nucleic acid
;       STRANDEDNESS:  single
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  cDNA
;     FEATURE:
;       NAME/KEY:  CDS
;       LOCATION:  1..6306
;     PUBLICATION INFORMATION:
;       AUTHORS:  COMPTON, DUANE A
;       AUTHORS:  SZILAK, ILLYA
;       AUTHORS:  CLEVELAND, DON W
;       TITLE:  PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR
;       TITLE:  PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
;       TITLE:  SEGREGATION OF PROTEINS AT MITOSIS
;       JOURNAL:  JOURNAL OF CELL BIOLOGY
;       VOLUME:  116
;       PAGES:  1395-1408
;       DATE:  MAR-1992
US-08-195-487-3

Alignment Scores:
Pred. No.:              3.2e-19        Length:          6306
Score:                  412.00         Matches:         372
Percent Similarity:     35.28%         Conservative:    256
Best Local Similarity:  20.90%         Mismatches:      663
Query Match:            5.10%          Indels:          493
DB:                     1              Gaps:            72

US-09-902-432-4 (1-1596) x US-08-195-487-3 (1-6306)

Qy       28 GlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspPro 47
Db     1050 GGCCACTCAGGAGTGGCTAGAGAAGCAGGCCCAGCTGGAGAAGGAGCTCAGCGCAGCCCT 1109

Qy       48 Ala-ThrLys------LeuProGlnLysAsn-----------GlyGlnLeuSerSerVa 61
Db     1110 GCAGGACAAGAAATGCCTTGAAGAGAAGAACGAAATCCTTCAGGGAAAACTTTCACAGCT 1169

Qy       61 l--------------------------AsnGlyValAlaGluGlnGlyAspVal----- 70
Db     1170 GGAAGAACACTTGTCCCAGCTGCAGGATAACCCACCCCAGGAGAAGGGCGAGGTGCTGGG 1229

Qy       71 ----------HisValGlnGluGluAsnGlnGluGly---------------------- 79
Db     1230 TGATGTCTTGCAGCTGGAAACCTTGAAGCAAGAGGCAGCCACTCTTGCTGCAAACAACAC 1289

Qy       80 ----------GlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAs 96
Db     1290 ACAGCTCCAAGCCAGGGTAGAGATGCTGGAGACTGAGCGAGGCCAGCAGGAAGCCAAGCT 1349

Qy       96 pValArgGluLysAspArgValGluGlu------------------------------ 105
Db     1350 GCTTGCTGAGCGGGGCCACTTCGAAGAAGAAAAGCAGCAGCTGTCTAGCCTGATCACTGA 1409

Qy      106 ------------MetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGl 121
Db     1410 CCTGCAGAGCTCCATCTCCAACCTCAGCCAGGCCAAGGAAGAGCTGGAGCAGGCCTCCCA 1469

Qy      121 nGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMe 141
Db     1470 GGCTCATGGGGCCCGGTTGACTGCCCAGGTGGCCTCTCTGACCTCTGAGCTCACCACACT 1529

Qy      141 tValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheVa 161
Db     1530 CAATGCCACCATCCAGCAACAGGATCAAGAACTG----------------------GC 1565

Qy      161 lGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLe 181
Db     1566 TGGCCTGAAGCAGCAGGCCAAAGAGAAGCAGGCCCAGCTAGCACAGACCCTCCAA----- 1620

Qy      181 uThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGl 201
Db     1621 ---------------------CAGCAAGAACAGGCCTCCCAGGGCCTC---CGCCACCA 1655

Qy      201 nGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGl 221
Db     1656 G---------GTGGAGCAGCTAAGCAGTAGCCTGAAGCAGAAGGAGCAGCAGTTGAAGGA 1706

Qy      221 nSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLe 241
Db     1707 GGTAGCGGAGAAGCAGGAGGCAACT------AGGCAGGACCATGCCCAGCAACTGGCCAC 1760

Qy      241 uGlnAlaGluSerAspGlnAlaAla------GluGluGluAlaLysAspGluGlyGluGl 259
Db     1761 TGCTGCAGAGGAGCGAGAGGCCTCCTTAAGGGAGCGGGATGCGGCTCTCAAGCAGCTGGA 1820

Qy      259 uLysGlnGluLysGluProThrLysSerProGlu------SerProSerSerProValAs 277
Db     1821 GGCACTGGAGAAGGAGAAGGCTGCCAAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGC 1880

Qy      277 nSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLy 297
Db     1881 TAATGAAGCCCGGGACAGTGCCCAGACCTCAGTGACACAG---------GCCCAGCGGGA 1931

Qy      297 sLysThrSerPheLysLysSerLysGluAsp---------AspLeuGluThrAlaGluLy 314
Db     1932 GAAGGCAGAGCTGAGCCGGAAGGTGGAGGAACTCCAGGCCTGTGTTGAGACAGCCCGCCA 1991

Qy      314 sArgLysGluGlnGluAlaGluLysValAspGluGluGlu---------LysGluLysTh 331
Db     1992 GGAACAGCATGAGGCCCAGGCCCAGGTTGCAGAGCTAGAGTTGCAGCTGCGGTCTGAGCA 2051

Qy      331 rGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAl 351
Db     2052 GCAAAAAGCAACTGAGAAAGAAAGGGTGGCCCAGGAGAAGGACCAG------CTCCAGGA 2105

Qy      351 aAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSe 371
Db     2106 GCAGCTCCAGGCCCTCAAAGAGTCCTTGAAGGTCACCAAGGGCAGCCTTGAA-------- 2157

Qy      371 rGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGl 391
Db     2158 -GAGGAGAAG------CGCAGGGCTGCAGATGCCCTGGAAGAGCAGCAGCGTTGTATCTC 2210

Qy      391 nGluValValAlaGluVal------------------HisValSerThrValGluLysTh 405
Db     2211 TGAGCTGAAGGCAGAGACCCGAAGCCTGGTGGAGCAGCATAAGCGGGAACGAAAGGAGCT 2270

Qy      405 rGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGl 425
Db     2271 GGAAGAAGAGAGGGCTGGGCGCAAGGGGCTGGAGGCTCGATTACTGCAGCTTGGGGAGGC 2330

Qy      425 yGluSerLeuProProGlu------------LysLeuAlaGluProGlnGluValProGl 441
```

```
                                  |||              :::|||||||||
Db   2331 CCATCAGGCTGAGACTGAAGTCCTGCGGCGGGAGCTGGCAGAGGCCATGGCTGCCCAGCA 2390

Qy    441 nGluAlaGlu---ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGl 460
             ||||||         |||:::|||:::|||                     |||
Db   2391 CACAGCTGAGAGTGAGTGTGAGCAGCTCGTCAAAGAAGTAGCTGCCTGGCGTGACGGGTA 2450

Qy    460 yAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGl 480
           :::   :::|||                   ||||||   |||:::    :::|||
Db   2451 TGAGGATAGCCAG---------------CAAGAGGAGGC-ACAGTATGGCGCCATGTTCC 2494

Qy    480 uGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySe 500
           |||         :::   |||
Db   2495 AGGAACAGCTGATGACTTTGAA-------------------------------------- 2516

Qy    500 rProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLy 520
                                                     |||          ::
Db   2517 --------------------------------------------GGAGGAATGTGAGA 2530

Qy    520 sGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGl 540
           :|||       |||      |||    :::      |||            |||
Db   2531 AGGCCCGCCAGGAGCTGCAGGAGGCAAAGGAGAAGGTGGCAGGCATAGAATCCCACAGCG 2590

Qy    540 uSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPr 560
           :::|||        |||:::        |||   ::::::   :::|||   :::||
Db   2591 AGCTCCA---------GATAAGCCGGCAGCAGAACAAACTAGCTGAGCTCCATGCCAACC 2641

Qy    560 oGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAla-- 579
           |    ::::::|||         ||||||   |||      |||      ||||||
Db   2642 TGGCCAGAGCACT------CCAGCAGGTCCAAGAGAAGGAAGTCAGGGCCCAGAAGCTTG 2695

Qy    580 ------------------------GluGluGly-ThrThrSerAspGlyGluLysLysA 591
                                  ::::::|||     ||||||           :::
Db   2696 CAGATGACCTCTCCACTCTGCAGGAAAAGATGGCTGCCACCAGCAAAGAGGTGGCCCGCT 2755

Qy    591 rgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValA 611
              |||   :::      |||         ::::::         ::::::   |||:
Db   2756 TGGAGACCTTGGTCCGCAAGGCAGGTGAGCAGCAGGAAACAGCCTCCCGGGAGTTAGTCA 2815

Qy    611 rgArgProSerGluSer---AspLysGluGluGlu---LeuGluLysValLysSerAlaT 629
           ::     |||:::    :::     |||::::::    |||    ||||||:::    :::
Db   2816 AGGAGCCTGCGAGGGCAGGAGACAGACAGCCCGAGTGGCTGGAAGAGCAACAGGGACGCC 2875

Qy    629 hrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyG 649
                   ||||||    :::    :::         |||:::   |||    :::   :::|||
Db   2876 AGTTCTGCAGCACACAGGCAGCGCTGCAGGCTATGGAGCGGGAGGCAGAGCAGATGGGCA 2935

Qy    649 luGlu---------------------------------------GlnLysProGluGlu- 655
           |||                                          |||:::   ||||||
Db   2936 ATGAGCTGGAACGGCTGCGGGCCGCGCTGATGGAGAGCCAGGGGCAGCAGCAGGAGGAGC 2995

Qy    655 ------------------------------------------------------------ 655

Db   2996 GTGGGCAGCAGGAAAGGGAGGTGGCGCGGCTGACCCAGGAGCGGGGCCGTGCCCAGGCTG 3055

Qy    656 -----------------------------------------------------------P 656

Db   3056 ACCTTGCCCTGGAGAAGGCGGCCAGAGCAGAGCTTGAGATGCGGCTGCAGAACGCCCTCA 3115

Qy    656 roLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerL 676
             ::::::||||||:::    :::             |||||||||        :::    |
Db   3116 ACGAGCAGCGTGTGGAGTTCGCTACCCTGCAAGAGGCACTGGCTCATGCCCTGACGGAAA 3175

Qy    676 ysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyG 696
           |||:::                      |||    |||    |||          :::     |
Db   3176 AGGAA------------------GGCAAGGACCAGGAGTT-GGCCAAGCTTCGTGGTCTG 3216

Qy    696 lyAspSer------HisArgAlaGluGluAlaSerLysAspLysGluAla------GlyT 712
           ||    |||               |||||||        ::::::||||||        ||||
Db   3217 GAGGCAGCCCAGATAAAAGAGCTGGAGGAACTTCGGCAAACCGTGAAGCAACTGAAGGAA 3276

Qy    712 hrAspAlaValProAla-SerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPro 731
           ||                 |||  :::     :::|||::::::     |||       ||||||     :::

Db   3277 CA------------GCTGGCTAAGAAAGAAAAGGAGCACGCATCTGGCTCAGGAGCCCAA 3324

Qy    732 GluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArg 751
               |||         :::|||      |||                |||:::      :::
Db   3325 TCTGAGGCTGCTGGCAGGACAGAGCCAACAGGCCCCAAGCTG---GAAGCACTGCGGGCA 3381

Qy    752 LeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGlu--------- 768
             |||:::          ::::::               |||     :::|||:::|||:::
Db   3382 GAGGTGAGCAAGCTGGAACAGCAATGCCAGAAGCAGCAGGAGCAGGCTGACAGCCTGGAA 3441

Qy    769 --------------------------------------AspSerSerValGluGlnLeu 775
                                                  ||||||::::::|||     |||
Db   3442 CGCAGCCTCGAGGCTGAGCGGGCCTCCCGGGCTGAGCGGGACAGTGCTCTGGAGACTCTG 3501

Qy    776 SerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIlePro 795
             ::::::|||          :::|||          |||
Db   3502 CAGGGCCAGTTAGAGGAGAAGGCCCAGGAGCTAGGGCACAGTCAGAGTGCCTTAGCCTCG 3561

Qy    796 GlyArgArgLysLysArgAla------AspGlyLysGlnGluGlnAlaThrValGluAsp 813
            :::|||:::          |||          |||:::    :::          ||||||
Db   3562 GCCCAACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAAGACCACAGCAAGGCTGAAGAT 3621

Qy    814 SerGlyProValGluIle------------AsnGluAspAspProAsnValProAlaVal 829
                     ::::::                     |||    :::               :::
Db   3622 GAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCTGAGAGGAAAAATAGCCTCATC 3681

Qy    830 ValProLeu---SerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsn 848
                |||      |||    :::    :::    |||:::    :::|||    :::|||
Db   3682 AGCAGCTTGGAGGAGGAGGTGTCCATCCTGAATCGCCAGGTCCTGGAGAAGGAGGGG--- 3738

Qy    849 ThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeu 868
             |||       :::|||         |||                |||    |||:::    |||
Db   3739 ---GAGAGCAAGGAGTTGAAGCGGCTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTG 3795

Qy    869 ValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGlu 888
             :::   :::                                                  :::
Db   3796 GAGGAGAGCTGCGCCTGCTGC--------------------------------AGGCAG 3822

Qy    889 ArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGlu 908
            |||   |||:::
Db   3823 AGACAGCCAGCA------------------------------------------------ 3834

Qy    909 AlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVal 928
                     |||    |||:::                ::::::
Db   3835 ---------ACAGTGCCAGAGCTGCAGAACGCAGCTCTGCTCTGCGGGAGG--------- 3876

Qy    929 LeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGlu--- 947
             :::                 |||::::::|||                |||    ||||||
Db   3877 ------AGGTGCAGAGCCTCCGGGAGGGAGGCTGAGAAACAGCGGGTGGCTTCAGAGAAC 3930

Qy    948 ------ValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArg 965
                  :::          ||||||:::|||          |||    |||       :::    |||:::
Db   3931 CTGCGGCAGGAGCTGACCTCACAGGCTGAGCGTGCGGAGGAGCTGGGCCAAGAATTGAAG 3990

Qy    966 Thr------GluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSer 983
                    |||:::          :::          |||          |||          :::
Db   3991 GCGTGGCAGGAGAAGTTCTTCCAGAAAGAGCAGGCCCTCTCCACCCTGCAGCTC------ 4044

Qy    984 AlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGlu 1003
                                                        |||     |||          |||
Db   4045 --------------------------------------GAGCACACCAGCACACAGGCC 4065

Qy   1004 ValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGln 1023
            :::    |||    :::|||          :::          :::|||    ||||||
Db   4066 CTGGTGAGTGAGCTGCTGCCAGCTAAGCACCTCTGCCAGCAGCTGCAGGCCGAGCAGGCC 4125

Qy   1024 AlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArg 1043
            |||          :::    :::||||||    :::                |||:::    |||
Db   4126 GCTGCCGAGAAACGCCACCGTGAGGAGCTGGAG------------CAGAGCAAGCAGGCC 4173

Qy   1044 ThrGlySerLysAlaLeuGlu--------------------------------------- 1050
               |||          |||
Db   4174 GCTGGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAGCGGGAGCTTGGGGAGCTGATTCCT 4233
```

```
Qy   1051 ---------LysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys--- 1066
Db   4234 CTGCGGCAGAAGGTGGCAGAGCAGGACCGAACAGCTCAGCAGCTGCGGGCAGAGAAGGCC 4293

Qy   1066 ------------------------------------------------------------ 1066
Db   4294 AGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTGCTGGCAGAGGAGAAC 4353

Qy   1066 ------------------------------------------------------------ 1066
Db   4354 CGGGGGCTGGGTGAGCGGGCCAACCTTGGCCGGCAGTTTCTGGAAGTGGAGTTGGACCAG 4413

Qy   1067 ------GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeu 1084
Db   4414 GCCCGGGAAAAGTATGTCCAAGAGTTGGCAGCCGTACGTGCTGATGCTGAGACCCGTCTG 4473

Qy   1085 AlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThr 1104
Db   4474 GCTGAGGTGCAGCGAGAAGCACAGAGCACTGCCCGGGAGCTGGAGGTG------ATGACT 4527

Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGln 1124
Db   4528 GCCAAGTATGAGGGTGCC------------------AAGGTCAAGGTCCTGGAGGAGAGG 4569

Qy   1125 AlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrPro 1144
Db   4570 CAGCGGTTCCAGGAAGAGAGGCAGAAACTCACTGCCCAGGTGGAAGAACTGAGTAAGAAA 4629

Qy   1145 LeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAsp------GluThrIleAspSer 1162
Db   4630 CTGGCTGACTCTGACCAAGCCAGCAAGGTGCAGCAGCAGAAGCTGAAGGCTGTCCAGGCT 4689

Qy   1163 Gln------AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
Db   4690 CAGGGAGGCGAGAGCCAGCAGGAGGCCCAGCGCTTCCAGGCCCAGCTGAATGAACTGCAA 4749

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSer------------------------ThrLeu 1192
Db   4750 GCCCAGTTGAGCCAGAAGGAGCAGGCAGCTGAGCACTATAAGCTGCAGATGGAGAAAGCC 4809

Qy   1193 ProAsnAsnValProAlaGlnGluGluHisGlyGluGlu---------------ProGly 1207
Db   4810 AAAACACATTATGATGCCAAGAAGCAGCAGAACCAAGAGCTGCAGGAGCAGCTGCGGAGC 4869

Qy   1208 ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAla 1227
Db   4870 CTGGAGCAGCTGCAGAAGGAAAACAAAGAGCTGCGAGCTGAAGCT--------------- 4914

Qy   1228 LysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu 1247
Db   4915 ---GAACGGCTGGGCCATGAGCTACAG---------CAGGCTGGGCTGAAGACCAAGGAG 4962

Qy   1248 GluGlnGluValPheValHis------SerGlyProAsnSerGlnLysAlaAlaAspVal 1265
Db   4963 GCTGAACAGACCTGCCGCCACCTTACTGCCCAGGTGCGCAGCCTGGAGGCACAGGTTGCC 5022

Qy   1266 ThrTyrAspSerGluValMetGlyValAlaGlyCysGln------------GluLysGlu 1281
Db   5023 CATGCAGACCAGCAGCTTCGAGACCTGGGCAAATTCCAGGTGGCAACTGATGCTTTAAAG 5082

Qy   1282 SerThrGluValGlnSer------------------------------------LeuSer 1289
Db   5083 AGCCGTGAGCCCCAGGCTAAGCCCCAGCTGGACTTGAGTATTGACAGCCTGGATCTGAGC 5142

Qy   1290 LeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGlu 1309
Db   5143 TGCGAGGAGGGGACCCCACTCAGTATCACCAGCAAGCTGCCTCGT---ACCCAGCCAGAC 5199

Qy   1310 GlnValSerGluGluGlyGluGlnGlu---------ThrAlaAlaProGluHisGluGly 1326
Db   5200 GGCACCAGCGTCCCTGGAGAACCAGCCTCACCTATCTCCCA-GCGCCTGCCCCCCAAGGT 5258

Qy   1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeu 1346
Db   5259 AGAATCCCTGGAGAGTCTCTACTTCACTCCCATCCCTGCTCGGAGTCAGGCCCCCCT--- 5315

Qy   1347 GlySerLeuGlyGlySerProSerLeuPro------------------------------ 1356
Db   5316 ---------GGAGAGCAGCCTGGACTCCCTGGGAGACGTCTTCCTGGACTCGGGTCGTAA 5366

Qy   1357 ------------------------------------------------AspGlnAspLys 1360
Db   5367 GACCCGCTCCGCTCGTCGGCGCACCACGCAGATCATCAACATCACCATGACCAAGAA--- 5423

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeu-AspThrThrValThrGlnThrAlaGl 1380
Db   5424 GCTAGATGTGGAAGAGCCAGACAGCGCCAACTCATCGTTCTACAGCACGCGGTCTGCTCC 5483

Qy   1380 uAlaValGluLysValIle-----------GluThrValValIleSerGluThrGlyGl 1396
Db   5484 TGCTTCCCAGGCTAGCCTGCGAGCCACCTCCTCTACTCAGTCTCTAGCTCGCCTGGGT-- 5541

Qy   1396 uSerProGluCysValGlyAlaHisLeuLeu---------------------------- 1406
Db   5542 -TCTCCCGATTATGGCAACTCAGCCCTGCTCAGCTTGCCTGGCTACCGCCCCACCACTCG 5600

Qy   1407 ----ProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAs 1425
Db   5601 CAGTTCTGCTCGTCGTTCCCAGGCCGG-GGTGTCCAGTGGGGCCCCTCCAGGAAGGAACA 5659

Qy   1425 pThrValPro-----------------------------------LeuGlyProGluSe 1433
Db   5660 GCTTCTACATGGGCACTTGCCAGGATGAGCCTGAGCAGCTGGATGACTGGAACCGCATTG 5719

Qy   1433 rGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisProAs 1453
Db   5720 CAGAGCTGCAGCAGCGCAATCGAGTGTGCCCCCACATCTGAAGACCTGCTATCCCCTGG 5779

Qy   1453 pLeuGlnGlyGluIleSerAlaSerGlnArgGluArg---SerGluGluGluAspLysPr 1472
Db   5780 AGTCCAGGCCTTCCCTGAGCCTGGGCACCATCACAGATGAGGAGATGAAAACTGGAGACC 5839

Qy   1472 oAsp--------------------------AlaGlyProAspAlaAspGly 1480
Db   5840 CCCAAGAGACCCTGCGCCGAGCCAGCATGCAGCCAATCCAGATAGCCGAGGG 5891
```

```
RESULT 12
PCT-US93-06160-3
; Sequence 3, Application PC/TUS9306160
;  GENERAL INFORMATION:
;    APPLICANT:
;    TITLE OF INVENTION:  NOVEL MALIGNANT CELL TYPE MARKERS OF THE
;    TITLE OF INVENTION:  INTERIOR NUCLEAR MATRIX
;    NUMBER OF SEQUENCES:  6
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  TESTA HURWITZ & THIBEAULT
;      STREET:  53 STATE STREET
;      CITY:  BOSTON
;      STATE:  MA
;      COUNTRY:  USA
;      ZIP:  02109
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  PCT/US93/06160
;      FILING DATE:  19930621
;      CLASSIFICATION:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  PITCHER ESQ, EDMUND R
;      REGISTRATION NUMBER:  27,829
;      REFERENCE/DOCKET NUMBER:  MTP-013
;    TELECOMMUNICATION INFORMATION:
```

```
;      TELEPHONE:  617/248-7000
;      TELEFAX:  617/248-7100
;  INFORMATION FOR SEQ ID NO:  3:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  6306 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1..6306
;    PUBLICATION INFORMATION:
;      AUTHORS:  COMPTON, DUANE A
;      AUTHORS:  SZILAK, ILLYA
;      AUTHORS:  CLEVELAND, DON W
;      TITLE:  PRIMARY STRUCTURE OF OF NUMA, AN INTRANUCLEAR
;      TITLE:  PROTEIN THAT DEFINES A NOVEL PATHWAY FOR
;      TITLE:  SEGREGATION OF PROTEINS AT MITOSIS
;      JOURNAL:  JOURNAL OF CELL BIOLOGY
;      VOLUME:  116
;      PAGES:  1395-1408
;      DATE:  MAR-1992
PCT-US93-06160-3

Alignment Scores:
Pred. No.:              3.2e-19        Length:        6306
Score:                  412.00         Matches:       372
Percent Similarity:     35.28%         Conservative:  256
Best Local Similarity:  20.90%         Mismatches:    663
Query Match:            5.10%          Indels:        493
DB:                     5              Gaps:          72

US-09-902-432-4 (1-1596) x PCT-US93-06160-3 (1-6306)

Qy      28 GlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspPro 47
              |||||          |||       :::|||   ||||||:::   |||           |||
Db    1050 GGCCACTCAGGAGTGGCTAGAGAAGCAGGCCCAGCTGGAGAAGGAGCTCAGCGCAGCCCT 1109

Qy      48 Ala-ThrLys------LeuProGlnLysAsn------------GlyGlnLeuSerSerVa 61
              |||    |||       |||   :::||||||             |||:::||||||    ::
Db    1110 GCAGGACAAGAAATGCCTTGAAGAGAAGAACGAAATCCTTCAGGGAAAACTTTCACAGCT 1169

Qy      61 l--------------------------AsnGlyValAlaGluGlnGlyAspVal----- 70
           :                             |||          |||:::|||:::|||
Db    1170 GGAAGAACACTTGTCCCAGCTGCAGGATAACCCACCCCAGGAGAAGGGCGAGGTGCTGGG 1229

Qy      71 ----------HisValGlnGluGluAsnGlnGluGly----------------------- 79
                     ::::::                  ||||||
Db    1230 TGATGTCTTGCAGCTGGAAACCTTGAAGCAAGAGGCAGCCACTCTTGCTGCAAACAACAC 1289

Qy      80 ----------GlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAs 96
                               |||:::::::::    :::    ||||||:::|||::::::
Db    1290 ACAGCTCCAAGCCAGGGTAGAGATGCTGGAGACTGAGCGAGGCCAGCAGGAAGCCAAGCT 1349

Qy      96 pValArgGluLysAspArgValGluGlu------------------------------- 105
            :::    |||:::           ||||||
Db    1350 GCTTGCTGAGCGGGGCCACTTCGAAGAAGAAAAGCAGCAGCTGTCTAGCCTGATCACTGA 1409

Qy     106 -------------MetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGl 121
                        ::::::       |||    |||   |||:::::::     :::      ||
Db    1410 CCTGCAGAGCTCCATCTCCAACCTCAGCCAGGCCAAGGAAGAGCTGGAGCAGGCCTCCCA 1469

Qy     121 nGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMe 141
           |         :::    :::          |||:::     :::      :::      ::
Db    1470 GGCTCATGGGGCCCGGTTGACTGCCCAGGTGGCCTCTCTGACCTCTGAGCTCACCACACT 1529

Qy     141 tValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheVa 161
           :          :::     |||       ::::::
Db    1530 CAATGCCACCATCCAGCAACAGGATCAAGAACTG-----------------------GC 1565

Qy     161 lGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLe 181
             |||    |||          |||:::    :::      :::     :::     |||:::|||
Db    1566 TGGCCTGAAGCAGCAGGCCAAAGAGAAGCAGGCCCAGCTAGCACAGACCCTCCAA----- 1620

Qy     181 uThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGl 201
                                 :::       :::||||||    |||             |||||
Db    1621 ---------------------CAGCAAGAACAGGCCTCCCAGGGCCTC---CGCCACCA 1655

Qy     201 nGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGl 221
           |             ||||||                          ||||||     :::||||||::
Db    1656 G---------GTGGAGCAGCTAAGCAGTAGCCTGAAGCAGAAGGAGCAGCAGTTGAAGGA 1706

Qy     221 nSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLe 241
           :          |||||||||||     |||          ::::::         :::     ::::::
Db    1707 GGTAGCGGAGAAGCAGGAGGCAACT------AGGCAGGACCATGCCCAGCAACTGGCCAC 1760

Qy     241 uGlnAlaGluSerAspGlnAlaAla------GluGluGluAlaLysAspGluGlyGluGl 259
                 ||||||          :::|||:::      |||    :::|||          :::    ||
Db    1761 TGCTGCAGAGGAGCGAGAGGCCTCCTTAAGGGAGCGGGATGCGGCTCTCAAGCAGCTGGA 1820

Qy     259 uLysGlnGluLysGluProThrLysSerProGlu------SerProSerSerProValAs 277
           |          |||||||||                |||                    |||
Db    1821 GGCACTGGAGAAGGAGAAGGCTGCCAAGCTGGAGATTCTGCAGCAGCAACTTCAGGTGGC 1880

Qy     277 nSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLy 297
             :::|||          |||    :::            |||                   |||::
Db    1881 TAATGAAGCCCGGGACAGTGCCCAGACCTCAGTGACACAG---------GCCCAGCGGGA 1931

Qy     297 sLysThrSerPheLysLysSerLysGluAsp---------AspLeuGluThrAlaGluLy 314
            :|||             :::        |||:::             :::||||||||    ::
Db    1932 GAAGGCAGAGCTGAGCCGGAAGGTGGAGGAACTCCAGGCCTGTGTTGAGACAGCCCGCCA 1991

Qy     314 sArgLysGluGlnGluAlaGluLysValAspGluGluGlu---------LysGluLysTh 331
           :     :::    :::          :::|||    |||   |||         :::     :::
Db    1992 GGAACAGCATGAGGCCCAGGCCCAGGTTGCAGAGCTAGAGTTGCAGCTGCGGTCTGAGCA 2051

Qy     331 rGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAl 351
             :::      |||:::|||::::::           |||::::::    ||||||       |||
Db    2052 GCAAAAAGCAACTGAGAAAGAAAGGGTGGCCCAGGAGAAGGACCAG------CTCCAGGA 2105

Qy     351 aAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSe 371
                 :::    ::::::          |||:::             |||    ||||||
Db    2106 GCAGCTCCAGGCCCTCAAAGAGTCCTTGAAGGTCACCAAGGGCAGCCTTGAA-------- 2157

Qy     371 rGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGl 391
             ||||||||           |||       :::        :::|||:::
Db    2158 -GAGGAGAAG------CGCAGGGCTGCAGATGCCCTGGAAGAGCAGCAGCGTTGTATCTC 2210

Qy     391 nGluValValAlaGluVal-----------------HisValSerThrValGluLysTh 405
             |||:::    ||||||                     |||               ::::::
Db    2211 TGAGCTGAAGGCAGAGACCCGAAGCCTGGTGGAGCAGCATAAGCGGGAACGAAAGGAGCT 2270

Qy     405 rGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGl 425
             |||||||||:::     |||             ||     ::::::        |||
Db    2271 GGAAGAAGAGAGGGCTGGGCGCAAGGGGCTGGAGGCTCGATTACTGCAGCTTGGGGAGGC 2330

Qy     425 yGluSerLeuProProGlu------------LysLeuAlaGluProGlnGluValProGl 441
                           |||             :::|||||||||
Db    2331 CCATCAGGCTGAGACTGAAGTCCTGCGGCGGGAGCTGGCAGAGGCCATGGCTGCCCAGCA 2390

Qy     441 nGluAlaGlu---ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGl 460
                 ||||||            |||:::|||:::|||                      |||
Db    2391 CACAGCTGAGAGTGAGTGTGAGCAGCTCGTCAAAGAAGTAGCTGCCTGGCGTGACGGGTA 2450

Qy     460 yAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGl 480
             :::    :::|||                      ||||||   |||:::     :::|||
Db    2451 TGAGGATAGCCAG--------------CAAGAGGAGGC-ACAGTATGGCGCCATGTTCC 2494

Qy     480 uGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySe 500
             |||          :::    |||
Db    2495 AGGAACAGCTGATGACTTTGAA-------------------------------------- 2516

Qy     500 rProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLy 520
                                                          |||        ::
Db    2517 --------------------------------------------GGAGGAATGTGAGA 2530
```

```
Qy    520 sGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGl 540
Db   2531 AGGCCCGCCAGGAGCTGCAGGAGGCAAAGGAGAAGGTGGCAGGCATAGAATCCCACAGCG 2590

Qy    540 uSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPr 560
Db   2591 AGCTCCA---------GATAAGCCGGCAGCAGAACAAACTAGCTGAGCTCCATGCCAACC 2641

Qy    560 oGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAla-- 579
Db   2642 TGGCCAGAGCACT------CCAGCAGGTCCAAGAGAAGGAAGTCAGGGCCCAGAAGCTTG 2695

Qy    580 ------------------------GluGluGly-ThrThrSerAspGlyGluLysLysA 591
Db   2696 CAGATGACCTCTCCACTCTGCAGGAAAAGATGGCTGCCACCAGCAAAGAGGTGGCCCGCT 2755

Qy    591 rgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValA 611
Db   2756 TGGAGACCTTGGTCCGCAAGGCAGGTGAGCAGCAGGAAACAGCCTCCCGGGAGTTAGTCA 2815

Qy    611 rgArgProSerGluSer---AspLysGluGluGlu---LeuGluLysValLysSerAlaT 629
Db   2816 AGGAGCCTGCGAGGGCAGGAGACAGACAGCCCGAGTGGCTGGAAGAGCAACAGGGACGCC 2875

Qy    629 hrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyG 649
Db   2876 AGTTCTGCAGCACACAGGCAGCGCTGCAGGCTATGGAGCGGGAGGCAGAGCAGATGGGCA 2935

Qy    649 luGlu--------------------------------------GlnLysProGluGlu- 655
Db   2936 ATGAGCTGGAACGGCTGCGGGCCGCGCTGATGGAGAGCCAGGGGCAGCAGCAGGAGGAGC 2995

Qy    655 ------------------------------------------------------------ 655
Db   2996 GTGGGCAGCAGGAAAGGGAGGTGGCGCGGCTGACCCAGGAGCGGGGCCGTGCCCAGGCTG 3055

Qy    656 -----------------------------------------------------------P 656
Db   3056 ACCTTGCCCTGGAGAAGGCGGCCAGAGCAGAGCTTGAGATGCGGCTGCAGAACGCCCTCA 3115

Qy    656 roLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerL 676
Db   3116 ACGAGCAGCGTGTGGAGTTCGCTACCCTGCAAGAGGCACTGGCTCATGCCCTGACGGAAA 3175

Qy    676 ysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyG 696
Db   3176 AGGAA-----------------GGCAAGGACCAGGAGTT-GGCCAAGCTTCGTGGTCTG 3216

Qy    696 lyAspSer------HisArgAlaGluGluAlaSerLysAspLysGluAla------GlyT 712
Db   3217 GAGGCAGCCCAGATAAAAGAGCTGGAGGAACTTCGGCAAACCGTGAAGCAACTGAAGGAA 3276

Qy    712 hrAspAlaValProAla-SerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPro 731
Db   3277 CA-----------GCTGGCTAAGAAAGAAAAGGAGCACGCATCTGGCTCAGGAGCCCAA 3324

Qy    732 GluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArg 751
Db   3325 TCTGAGGCTGCTGGCAGGACAGAGCCAACAGGCCCCAAGCTG---GAAGCACTGCGGGCA 3381

Qy    752 LeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGlu--------- 768
Db   3382 GAGGTGAGCAAGCTGGAACAGCAATGCCAGAAGCAGCAGGAGCAGGCTGACAGCCTGGAA 3441

Qy    769 --------------------------------------AspSerSerValGluGlnLeu 775
Db   3442 CGCAGCCTCGAGGCTGAGCGGGCCTCCCGGGCTGAGCGGGACAGTGCTCTGGAGACTCTG 3501

Qy    776 SerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIlePro 795
Db   3502 CAGGGCCAGTTAGAGGAGAAGGCCCAGGAGCTAGGGCACAGTCAGAGTGCCTTAGCCTCG 3561

Qy    796 GlyArgArgLysLysArgAla------AspGlyLysGlnGluGlnAlaThrValGluAsp 813
Db   3562 GCCCAACGGGAGTTGGCTGCCTTCCGCACCAAGGTACAAGACCACAGCAAGGCTGAAGAT 3621

Qy    814 SerGlyProValGluIle------------AsnGluAspAspProAsnValProAlaVal 829
Db   3622 GAGTGGAAGGCCCAGGTGGCCCGGGGCCGGCAAGAGGCTGAGGAAAAATAGCCTCATC 3681

Qy    830 ValProLeu---SerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsn 848
Db   3682 AGCAGCTTGGAGGAGGAGGTGTCCATCCTGAATCGCCAGGTCCTGGAGAAGGAGGGG--- 3738

Qy    849 ThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeu 868
Db   3739 ---GAGAGCAAGGAGTTGAAGCGGCTGGTGATGGCCGAGTCAGAGAAGAGCCAGAAGCTG 3795

Qy    869 ValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGlu 888
Db   3796 GAGGAGAGCTGCGCCTGCTGC---------------------------------AGGCAG 3822

Qy    889 ArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGlu 908
Db   3823 AGACAGCCAGCA------------------------------------------------ 3834

Qy    909 AlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVal 928
Db   3835 ---------ACAGTGCCAGAGCTGCAGAACGCAGCTCTGCTCTGCGGGAGG--------- 3876

Qy    929 LeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGlu--- 947
Db   3877 ------AGGTGCAGAGCCTCCGGGAGGGAGGCTGAGAAACAGCGGGTGGCTTCAGAGAAC 3930

Qy    948 ------ValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArg 965
Db   3931 CTGCGGCAGGAGCTGACCTCACAGGCTGAGCGTGCGGAGGAGCTGGGCCAAGAATTGAAG 3990

Qy    966 Thr------GluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSer 983
Db   3991 GCGTGGCAGGAGAAGTTCTTCCAGAAAGAGCAGGCCCTCTCCACCCTGCAGCTC------ 4044

Qy    984 AlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGlu 1003
Db   4045 ---------------------------------------GAGCACACCAGCACACAGGCC 4065

Qy   1004 ValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGln 1023
Db   4066 CTGGTGAGTGAGCTGCTGCCAGCTAAGCACCTCTGCCAGCAGCTGCAGGCCGAGCAGGCC 4125

Qy   1024 AlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArg 1043
Db   4126 GCTGCCGAGAAACGCCACCGTGAGGAGCTGGAG------------CAGAGCAAGCAGGCC 4173

Qy   1044 ThrGlySerLysAlaLeuGlu--------------------------------------- 1050
Db   4174 GCTGGGGGACTGCGGGCAGAGCTGCTGCGGGCCCAGCGGGAGCTTGGGGAGCTGATTCCT 4233

Qy   1051 ---------LysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys--- 1066
Db   4234 CTGCGGCAGAAGGTGGCAGAGCAGGACCGAACAGCTCAGCAGCTGCGGGCAGAGAAGGCC 4293

Qy   1066 ------------------------------------------------------------ 1066
Db   4294 AGCTATGCAGAGCAGCTGAGCATGCTGAAGAAGGCGCATGGCCTGCTGGCAGAGGAGAAC 4353

Qy   1066 ------------------------------------------------------------ 1066
Db   4354 CGGGGGCTGGGTGAGCGGGCCAACCTTGGCCGGCAGTTTCTGGAAGTGGAGTTGGACCAG 4413

Qy   1067 ------GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeu 1084
Db   4414 GCCCGGGAAAAGTATGTCCAAGAGTTGGCAGCCCGTACGTGCTGATGCTGAGACCCGTCTG 4473

Qy   1085 AlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThr 1104
```

```
               |||:::                   |||:::|||             ||||||||       :::|||
Db    4474 GCTGAGGTGCAGCGAGAAGCACAGAGCACTGCCCGGGAGCTGGAGGTG------ATGACT 4527

Qy    1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGln 1124
           |||        :::                              |||::::::   |||   |||:::
Db    4528 GCCAAGTATGAGGGTGCC------------------AAGGTCAAGGTCCTGGAGGAGAGG 4569

Qy    1125 AlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrPro 1144
                          |||       :::   ||||||                   |||
Db    4570 CAGCGGTTCCAGGAAGAGAGGCAGAAACTCACTGCCCAGGTGGAAGAACTGAGTAAGAAA 4629

Qy    1145 LeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAsp------GluThrIleAspSer 1162
           |||||||||||||||   |||          ||||||         :::   :::   :::
Db    4630 CTGGCTGACTCTGACCAAGCCAGCAAGGTGCAGCAGCAGAAGCTGAAGGCTGTCCAGGCT 4689

Qy    1163 Gln------AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
           |||         :::|||:::       |||          |||:::|||:::   |||     :::
Db    4690 CAGGGAGGCGAGAGCCAGCAGGAGGCCCAGCGCTTCCAGGCCCAGCTGAATGAACTGCAA 4749

Qy    1181 AlaAlaThrAlaGlnLysGluGluProSer------------------------ThrLeu 1192
           |||        :::|||||||||:::   :::
Db    4750 GCCCAGTTGAGCCAGAAGGAGCAGGCAGCTGAGCACTATAAGCTGCAGATGGAGAAAGCC 4809

Qy    1193 ProAsnAsnValProAlaGlnGluGluHisGlyGluGlu---------------ProGly 1207
              :::        |||:::::::::      :::|||
Db    4810 AAAACACATTATGATGCCAAGAAGCAGCAGAACCAAGAGCTGCAGGAGCAGCTGCGGAGC 4869

Qy    1208 ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAla 1227
              :::      |||:::          :::||||||   |||   |||
Db    4870 CTGGAGCAGCTGCAGAAGGAAAACAAAGAGCTGCGAGCTGAAGCT--------------- 4914

Qy    1228 LysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu 1247
                    :::|||   |||   :::                |||   |||   ||||||
Db    4915 ---GAACGGCTGGGCCATGAGCTACAG---------CAGGCTGGGCTGAAGACCAAGGAG 4962

Qy    1248 GluGlnGluValPheValHis------SerGlyProAsnSerGlnLysAlaAlaAspVal 1265
              ::::::         |||      :::            |||   :::|||
Db    4963 GCTGAACAGACCTGCCGCCACCTTACTGCCCAGGTGCGCAGCCTGGAGGCACAGGTTGCC 5022

Qy    1266 ThrTyrAspSerGluValMetGlyValAlaGlyCysGln------------GluLysGlu 1281
                 |||   ::::::      :::         |||                  :::
Db    5023 CATGCAGACCAGCAGCTTCGAGACCTGGGCAAATTCCAGGTGGCAACTGATGCTTTAAAG 5082

Qy    1282 SerThrGluValGlnSer------------------------------------LeuSer 1289
           |||   |||   |||:::                                     ||||||
Db    5083 AGCCGTGAGCCCCAGGCTAAGCCCCAGCTGGACTTGAGTATTGACAGCCTGGATCTGAGC 5142

Qy    1290 LeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThrLysProGlu 1309
                 |||||||||         :::      |||      |||   |||:::|||:::
Db    5143 TGCGAGGAGGGGACCCCACTCAGTATCACCAGCAAGCTGCCTCGT---ACCCAGCCAGAC 5199

Qy    1310 GlnValSerGluGluGlyGluGlnGlu---------ThrAlaAlaProGluHisGluGly 1326
                 |||      ||||||                :::   ||||||      :::|||
Db    5200 GGCACCAGCGTCCCTGGAGAACCAGCCTCACCTATCTCCCA-GCGCCTGCCCCCCAAGGT 5258

Qy    1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeu 1346
                             |||               |||   ||||||   |||
Db    5259 AGAATCCCTGGAGAGTCTCTACTTCACTCCCATCCCTGCTCGGAGTCAGGCCCCCCT--- 5315

Qy    1347 GlySerLeuGlyGlySerProSerLeuPro------------------------------ 1356
                    |||         |||   ||||||
Db    5316 ---------GGAGAGCAGCCTGGACTCCCTGGGAGACGTCTTCCTGGACTCGGGTCGTAA 5366

Qy    1357 ------------------------------------------------AspGlnAspLys 1360
                                                           ||||||:::
Db    5367 GACCCGCTCCGCTCGTCGGCGCACCACGCAGATCATCAACATCACCATGACCAAGAA--- 5423

Qy    1361 AlaGlyCysIleGluValGlnValGlnSerLeu-AspThrThrValThrGlnThrAlaGl 1380
           |||   |||         :::   :::   |||                |||::::::|||
Db    5424 GCTAGATGTGGAAGAGCCAGACAGCGCCAACTCATCGTTCTACAGCACGCGGTCTGCTCC 5483

Qy    1380 uAlaValGluLysValIle-----------GluThrValValIleSerGluThrGlyGl 1396
            |||   :::      :::                |||       ::::::      |||
Db    5484 TGCTTCCCAGGCTAGCCTGCGAGCCACCTCCTCTACTCAGTCTCTAGCTCGCCTGGGT-- 5541

Qy    1396 uSerProGluCysValGlyAlaHisLeuLeu----------------------------- 1406
               ||||||:::          :::   ||||||
Db    5542 -TCTCCCGATTATGGCAACTCAGCCCTGCTCAGCTTGCCTGGCTACCGCCCCACCACTCG 5600

Qy    1407 ----ProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAs 1425
                  |||    :::|||   |||   |||      |||             |||
Db    5601 CAGTTCTGCTCGTCGTTCCCAGGCCGG-GGTGTCCAGTGGGCCCCTCCAGGAAGGAACA 5659

Qy    1425 pThrValPro-----------------------------------LeuGlyProGluSe 1433
              :::                                          |||   |||
Db    5660 GCTTCTACATGGGCACTTGCCAGGATGAGCCTGAGCAGCTGGATGACTGGAACCGCATTG 5719

Qy    1433 rGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisProAs 1453
              :::|||   :::                |||   |||      |||      |||   |||
Db    5720 CAGAGCTGCAGCAGCGCAATCGAGTGTGCCCCCACATCTGAAGACCTGCTATCCCCTGG 5779

Qy    1453 pLeuGlnGlyGluIleSerAlaSerGlnArgGluArg---SerGluGluGluAspLysPr 1472
              :::|||                              |||      :::|||      :::||
Db    5780 AGTCCAGGCCTTCCCTGAGCCTGGGCACCATCACAGATGAGGAGATGAAAACTGGAGACC 5839

Qy    1472 oAsp--------------------------AlaGlyProAspAlaAspGly 1480
              |                             |||   ||||||:::   |||
Db    5840 CCCAAGAGACCCTGCGCCGAGCCAGCATGCAGCCAATCCAGATAGCCGAGGG 5891
```

RESULT 13
US-08-931-999-4
; Sequence 4, Application US/08931999
; Patent No. 6043219
; GENERAL INFORMATION:
; APPLICANT: Iandolo, John J.
; APPLICANT: Crupper, Scott S.
; TITLE OF INVENTION: Broad Spectrum Chemotherapeutic Peptide
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hovey, Williams, Timmons & Collins
; STREET: 2405 Grand Boulevard, Suite 400
; CITY: Kansas City
; STATE: Missouri
; COUNTRY: U.S.A.
; ZIP: 64108
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,999
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/710,561
; FILING DATE: 19-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Collins, John M.
; REGISTRATION NUMBER: 26,262
; REFERENCE/DOCKET NUMBER: 25043-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 816/474-9050
; TELEFAX: 816/474-9057
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6755 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Staphylococcus aureus
; STRAIN: UT0007

US-08-931-999-4

```
Alignment Scores:
Pred. No.:                  4.46e-19      Length:         6755
Score:                      410.50        Matches:        319
Percent Similarity:         36.16%        Conservative:   218
Best Local Similarity:      21.48%        Mismatches:     643
Query Match:                5.08%         Indels:         312
DB:                         3             Gaps:           55

US-09-902-432-4 (1-1596) x US-08-931-999-4 (1-6755)

Qy    169 LysAspLysAsnGluLysSerAspThrValGlnLeuLeu------ThrValLysLysAsp 186
Db     15 AGAGAAGAAAACGAAAGAGAGGAAACAGAACAGACAGCAGAAAACGAGAGAAAGAAGCAG 74
Qy    187 GluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGlu 206
Db     75 CGGAACACCACCAGGGGGAGCGCCACCGGGGGAAGGAACAACCAAACAGCCAACAACGCG 134
Qy    207 ThrAlaValGlyGluSerAlaSer---------LysGluSerGluLeuLysGlnSerThr 223
Db    135 ACACGCGCGCAGGAAGGGAGAAGCAAAGGACACAAGAAAGAAGAACCAGAACAAAAAACA 194
Qy    224 GluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGlnAla 243
Db    195 GAAAAGAACAAACCA------AAAGAAGAAGAAACGAACCCAGAA--------------- 233
Qy    244 GluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLys 263
Db    234 ------------------------------AAAGAAGAAGGACAAGAGAAGGAAGAAGAC 263
Qy    264 GluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSerSer 283
Db    264 CACAA-AGCGAGAAAAGCGAAGGGCGGAGACAACGGGAACAAGCAAAAAAAACAAAACCA 322
Qy    284 PheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLys 303
Db    323 GCGAAAAAAAAGCGACGCGAAACGCAAGCCGCA------GAAAGAAGCGCAAAGAAAAAA 376
Qy    304 SerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGlu------------ 319
Db    377 AAAAAAGAAGACAACACAGAAGACAACACAAAAAGAAGAAAAAAACAGGACACAAACGAA 436
Qy    320 -----------------AlaGluLysValAspGluGluGluLysGluLysThrGluPro 333
Db    437 AGGGGGGGACGAGAAAAAAAAAAGAGCAAAACGAAGAAAACCAGAAAAAAACAAAACCG 496
Qy    334 AlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db    497 AAGAAAGAAAGAAAGAAACCA--------------------------------------- 517
Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db    518 ---AAGCCAAGAAGACCA-----------------------------AAAAGGGAACGC 544
Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGluVal 393
Db    545 AAAACGGCGCCA------------------GAAAAGAAAAAAAGCAAAAGCCAACCACAG 586
Qy    394 ValAlaGluValHisValSerThrValGluLysThrGluGlu-GluGlnGlyGlyGlyGl 413
Db    587 GAGAAGGAACAAAAAAAAAAAGAAGCAGAGCAAACACAAAGGAGAAGGGGGAGGCACCAA 646
Qy    413 yGluAlaGluGlyGlyValValValGluGly----------------------------- 423
Db    647 CGAAAAAAAAAAGAACAAAAAAGAAAGAAGCCGGAAAGCGCAAGAAAAACGAAGCAAACA 706
Qy    424 -ThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAl 443
Db    707 AAACGGACAAAACGGCCCACCAGAAAAA------CACACCCAAAAAAAGCGACAGAAAAA 760
Qy    443 aGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisTh 463
Db    761 AAAAAAAGCAAAAAAGAAAACGAAGGAACAGGAA-----------------CCCCAGCA 802
Qy    463 rGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleVa 483
Db    803 ACAAAAAACAGACGGAAAAGGACGGAAAAAAACA---AAAACACACCCCCAA-------- 851
Qy    483 lSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLy 503
Db    852 ---------------------AACAAAGAAGGCACCAAAAAAAAAGGCAAACACCAAAA 889
Qy    503 sLysLeuPheSer--------------SerSerGlyLeuLysLysLeuSerGlyLysLy 518
Db    890 AAAAGGGCACGCAAGAGAGAGAAAAAAGGGGAAAGGAAAAAAACCCAAGGAAGGGACCAA 949
Qy    518 sGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHi 538
Db    950 AAAGAAGGAAAAAAAAAAAAAAACGGAGAAACAGGAAAAAAAGGAAAACAAACAGAGAAG 1009
Qy    538 sThrGluSerProGluSerAlaAspGlu---GlnLysGlyGluSerSerAlaSerSerPr 557
Db   1010 GACGAACAAAACGAAACACGGACCCGAACCACAAAAAAAAGAAGCCAAGGCCCGAGGAAA 1069
Qy    557 oGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGl 577
Db   1070 AGAGACAAGAAAAAAAAGCACA--------------------GAGGACACAGCGGAAGG 1108
Qy    577 yGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTr 597
Db   1109 AGAACCCACAAAGAACAAAAAGCAGAGAACAGAAAAAAAG---CACGGACCAAACCCCAA 1165
Qy    597 pAlaSerPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAs 617
Db   1166 AGAAGAAAGAAAAAAGAACGACAGAAGACAAAAGACAAAGAGAAAAAAAAGCGGGCAAAA 1225
Qy    617 pLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrVa 637
Db   1226 AGCAGACCAAGAGACGAACAAAAGAAAG------------------------------ 1253
Qy    637 lSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLy 657
Db   1254 ----GAAAAAGAGGACGAAAGAAAAAAAACAAGAGCACACCAAAGAACAGACGAAAG-AA 1308
Qy    657 sArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLy 677
Db   1309 AAGAAGACCGGAAACCAGCCAAAACGGAAGGACCGCCCAAAGGAAGGGA---GGACCCAG 1365
Qy    677 sArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAs 697
Db   1366 GAGAGGCAGAAAAGGG---------------CAGAACGGCCCCCAGCAAACGGAAAACAA 1410
Qy    697 pSerHisArgAlaGluGluAla---SerLysAspLysGluAlaGlyThrAspAlaValPr 716
Db   1411 AAAAAACCAAAAGGAAAGAGGCAGAGGAAAAGACAAAAAGAGGAGAAAACAACGACAAGA 1470
Qy    716 oAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPro------GluProAl 734
Db   1471 AAACAGCGGGAAGGAGCCGGAAGCAGGCAAAGGAGCCAACGCGGAAAGACACGAAAACAA 1530
Qy    734 aGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValTh 754
Db   1531 AGGAAAAAACAGCAAAAGCGAAAGAGAAGAAAAAAGAAAGAAGAGAAAGGCGACGAAGAC 1590
Qy    754 rProArgLysLysSerLysSerLysLeuGluGlu---------LysAlaGluAspSerSe 771
Db   1591 AGGGAGAGCGGAGCAAAAGAGAAAAAACCAAGAAGAACCGACGAAAACAAAAAGAAAAGG 1650
Qy    771 rValGluGlnLeuSerThrGluIleGluPro---SerArgGluGluSerTrpValSerIl 790
Db   1651 AAAGAAGCAAAAAAAAAAAAGAAAAACACCCGGAAGCCAGGAAAAGCCAGAGAGCAAAA 1710
Qy    790 eLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaTh 810
```

```
Db   1711 AAAAAACAAAAGCAGAAAAAGAAGAAAAAGAAGAACAAAAAACAAACAAAACAAGGCAAA 1770
Qy    810 rValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValVa 830
Db   1771 AAAGAAAAAGGCG------------------GAAGACAAAAAAAACACGCGGGCAAGGAC 1812
Qy    830 lProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGl 850
Db   1813 CGGACACAGCAAA---CGAAGGAAAAAGCGGAACAAAAGAGAAGAAGAAAAAAGCAAAGA 1869
Qy    850 uLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHi 870
Db   1870 AAAACCACGA---------------------GCCGAACCGCAGGAGAGG----------- 1897
Qy    870 sThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSe 890
Db   1898 ----------------------GAGACGAAAAGAAGCAAAGACGAAAAAGAAACACGAAA 1935
Qy    890 rProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr------AlaGlyGl 908
Db   1936 A-----------GAAAAGAAAAAACACGAGCCGCGAAAGAAAACAAAGAGGAGAGGAGA 1983
Qy    908 uAlaMetProProValGluGluValThrGluLysAspIleIleAlaGlu---GluThrPr 927
Db   1984 AACACAGCCCAGGAAAAGACAGGAAAACCGGAGAAACAAGAAAGCAGAGGGAAAGAGCCC 2043
Qy    927 oValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGl 947
Db   2044 CCCAACGAGGGAAAGGCAAAGAAAAAAAAAAAAAAGAGCCACCACACAAAGCCAAACAGA 2103
Qy    947 uValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGl 967
Db   2104 AGCGGAGGGGGAAAAGAACGAAAAGAAGGCACCAGAACAGAAAAAAAAGAACCAGGAAGA 2163
Qy    967 uGluValThrGluAlaSerGly-------------------------------AlaGl 976
Db   2164 GAGGAGGGAAAACAGAAAAGGGGGAAGAAACAAAGGAAGAAAAGGAAAGAAAAAAGCCGA 2223
Qy    976 uGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGl 996
Db   2224 AGGAACCGAAAGCCGGAAAGGCGCA------------ACGGAGAGCAAAGAAAAAACGAG 2271
Qy    996 uGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluAr 1016
Db   2272 AAGGACCACAGAGGG-AGAAAAAAAAAGACA------------GAAAAGCAAGAAGAACG 2318
Qy   1016 gGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValPr 1036
Db   2319 GAAGACCCAAAGAAACAAAAACGAAACAACGGACAAAAAAAAA----------------- 2361
Qy   1036 oAlaThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGl 1056
Db   2362 ----ACAAAGGCCAAACAAAGAAAGAAAGAAAAAGCACGAGGGAAG-------------- 2403
Qy   1056 uGluAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProVa 1076
Db   2404 -CAAACAACAGCAACCCAAGCCACACAAAAAAAAGACAGCACAAA-CCAAAAACAAAAGA 2461
Qy   1076 lGln------GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaTh 1094
Db   2462 ACAACCAAGAAAAGCAGGCGCAGAGAAA---CGACAACAAAGAAAAAAAGGAAGAAGGGC 2518
Qy   1094 rProGluSerLeuGluValProGluValThrAlaAspValAspHisValAlaThrCysGl 1114
Db   2519 AAACGAAAGCCGCAAAAACCAGCGCCGCCCAGGAGAAAACGAAGAAAAGCAAACCGCACA 2578
Qy   1114 nValIleLysLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLe 1134
Db   2579 A-----------------------GAAAAGGCCCACAGGACGCACAGAAGAGGCACAAG 2614
Qy   1134 uThrAspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyTh 1154
Db   2615 GAGAAACCACAAGCAAAAAAACGAAACAGCAAGCAAAAACACAAACCAAAGGACAGAGAG 2674
Qy   1154 rGlnGlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArg----- 1172
Db   2675 ACAGAAGGAAGAGGCCAAAGACCAACAGCAGACAAAAAGGAAGAAGGGGAACAGAAACGA 2734
Qy   1173 ------------------------GlnSerGlnValThrGluGluAlaAlaThrAl 1184
Db   2735 CAAGAGAGGGCAACAGGACAAAAACGAAAGCAGGAAAAAAGAAGAAAACAAGAGGGGGC 2794
Qy   1184 aGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGl 1204
Db   2795 AAAGAACGGCCAAACGCCCACGCAAACAAAAAACCAAGGGGAAAGACAGGACAGGCGAGC 2854
Qy   1204 uGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValPr 1224
Db   2855 AGACAGCGGCAGGCAC---------------AAACAACAAACACACGAAGCGACGGCAAC 2899
Qy   1224 oValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLy 1244
Db   2900 GAAAAAGAAACAGAAGAAGAAACCCGAGGAGAACGAAAAGCAGACAAAGCAGCAGGAGAG 2959
Qy   1244 sValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAs 1264
Db   2960 AGAAAAAGAAGCACAGACACACCAGAAAGCCCAGCAAAAAACAAAAAGAAAAAGAGGGGA 3019
Qy   1264 pValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGl 1284
Db   3020 CACAAGCAGGGGCAGC------------------GGGACAGGAAAACACAAGGAAACAAG 3061
Qy   1284 uValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysAr 1304
Db   3062 ACCCGAACCCAGGGCAGCAAAGGGGGGAGAGAA-AACAAAAAAACAGAAAAAGACACAAG 3120
Qy   1304 gGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHi 1324
Db   3121 ACAA------CCACAAGAAGCAACCCAAAAGAACAAGGAAAAGAAGGGGGAACACAAGAA 3174
Qy   1324 sGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArg---Gl 1343
Db   3175 ACGAAAAACAAACGGCCAA------------------CAGCCAAAACAAAAGCGCCAGGG 3216
Qy   1343 yLysAla------LeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAl 1361
Db   3217 AAAGACAGAGCGAAAGGGGCAAGCAGGAGGCAAC------------------GACAAA-- 3256
Qy   1361 aGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAl 1381
Db   3256 ------------------------------------------------------------ 3256
Qy   1381 aValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCysVa 1401
Db   3257 ---------------------------------GAAACAAAGCAAAAGCCGCAA----- 3277
Qy   1401 lGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGl 1421
Db   3278 ----------------CCCAAAAAGAGGAGAAGCACC----------CACGA-AAGGCGCA 3311
Qy   1421 nHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIl 1441
Db   3312 GCACAGAAGACGAACACGAGAAAAAGCACCAAAACGAAGCCACGAGCAC----------- 3360
Qy   1441 eValThrProAlaProGluSerThrLeuHisProAsp------LeuGlnGlyGluIleSe 1459
Db   3361 ----------------GAAAGCAGAAAAAAGGAGGACAGGAGAAAGGAAGGGCGCACAAA 3404
Qy   1459 rAlaSerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAs 1479
Db   3405 GAAGAAAAAAAAGGAAAGAAAACAAAGCGAAAAAAAGGAGGGGAGGAGAAAAGAAAAAGA 3464
Qy   1479 pGlyLysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLe 1499
Db   3465 GAACAAAAAACACCGAGCAGAAGAAAAAAAAAAAGGGAAAAAAAAAACAAAAAAAAAAAAC 3524
```

```
Qy   1499 uGluSerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAl 1519
            |||   |||:::   :::              ||||||        |||
Db   3525 AGAAGACAAGACGAGGAGAAAAAAAAAAAGAACAAG-CAGACAACGAAGAAACAAGCGGA 3583

Qy   1519 aArgThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCy 1539
                |||:::          :::              :::|||   |||    |||
Db   3584 AAACACGAAAAAACGAGAACAGAGAGAAAAAAAAAAAACCCAAGCACAACCGCAAGCGAC 3643

Qy   1539 sArgLeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLy 1559
             :::      :::   :::||||||||         |||   |||:::   |||   ||
Db   3644 CAAAAAAAAAACGCCGAAACCAAACAGG------AAAAAACCAAAAGAGAGAAAGAAAAA 3697

Qy   1559 sHisProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAl 1579
          |              :::   |||         :::                |||
Db   3698 AAAAAAACGGAGCAAAAAGCGCAGAAAAACGGAG---------------GAAACAAGAAA 3742

Qy   1579 aGlnProArgLys 1583
             :::|||:::|||
Db   3743 AAAACCAAAAAAA 3755
```

RESULT 14
US-07-853-913-1
; Sequence 1, Application US/07853913
; Patent No. 5338839
; GENERAL INFORMATION:
;   APPLICANT: McKay, Ronald D.G.
;   APPLICANT: Lendahl, Urban
;   TITLE OF INVENTION: Nestin Expression As An Indicator of
;   TITLE OF INVENTION: Neuroepithelial Tumors
;   NUMBER OF SEQUENCES: 4
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;   STREET: Two Militia Drive
;   CITY: Lexington
;   STATE: Massachusetts
;   COUNTRY: U.S.A.
;   ZIP: 02173
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/853,913
;   FILING DATE: 19920319
;   CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/660,412
;   FILING DATE: 22-FEB-1991
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/603,803
;   FILING DATE: 25-OCT-1990
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/201,762
;   FILING DATE: 02-JUN-1988
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 07/180,548
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Granahan, Patricia
;   REGISTRATION NUMBER: 32,227
;   REFERENCE/DOCKET NUMBER: MIT-4641AAAA
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 617-861-6240
;   TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 11236 base pairs
;   TYPE: NUCLEIC ACID
;   STRANDEDNESS: double
;   TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-07-853-913-1

```
Alignment Scores:
Pred. No.:                 3.23e-17      Length:          11236
Score:                     388.00        Matches:         340
Percent Similarity:        37.54%        Conservative:    261
Best Local Similarity:     21.24%        Mismatches:      588
Query Match:               4.81%         Indels:          416
DB:                        1             Gaps:            82

US-09-902-432-4 (1-1596) x US-07-853-913-1 (1-11236)

Qy     13 GluGlnProAlaGlySerAspThr---------ProSerGluLeuVal------------ 25
          |||:::|||   |||                 |||   :::|||
Db   6721 GAGGAACCTGGGGGCAAGCAGCAGGGTCACTTCCCTGATGATCTGACCTCCTTAGCCACA 6780

Qy     26 ---LeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAsp 44
             |||:::   |||   |||      ||||||            |||:::   ::::::
Db   6781 AACCTCAACCCTCACCACCCTACTTTAGAGGCTAAAGAT------GGAGAATCCAGTGAG 6834

Qy     45 AlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAsnGlyVal 64
          :::      ::::::         :::      ||||||:::      :::      |||
Db   6835 TCTAGAGTTTCTAGCATATTCCAGGAAGATGAGGGGCAAATCTGGGAACTG------GTA 6888

Qy     65 AlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln------GluGlu 82
             ::::::   |||:::   |||:::   ||||||         |||      :::|||
Db   6889 GAGAAAGAAGCAGATATAGAGGTAAAAGTAGAAAACAGCTCAGCCCAGAAAACACAAGAA 6948

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
                :::|||   :::            |||   :::|||:::         :::|||:::
Db   6949 AGTGGTCTGGACACAGAAGAAACCCAG---GATTCCCAGGGACCTTTGCAGAAGGAAACA 7005

Qy    103 ValGluGluMetAlaAlaAsn---------SerThrAlaValGluAspIleThrLysAsp 119
          ::::::   :::                  |||      :::::::::
Db   7006 CTGAAGGCTCTAGGAGAGGAGCCACTGATGTCTCTGAAAATCCAGAACTATGAGACAGCA 7065

Qy    120 GlyGlnGlu---------------------------------------GluThrSerGlu 126
          |||:::|||                                       |||   |||
Db   7066 GGGAAAGAGAATTGCAATTCTTCTACAGAAGGCCACCTGGGAACACTAGAAGGCCCAGAA 7125

Qy    127 IleIleGluGlnIlePro---------AlaSerGluAsnAsnValGluGluMetValGln 143
                :::|||||||||               |||   |||||||||
Db   7126 AAAGAAAAGCAAATACCACTAAAGTCTTTAGAAGAAAAGAATGTAGAG------------ 7173

Qy    144 ProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGlyPhe 163
                :::         :::   |||         :::   :::   :::|||
Db   7174 ------TCAGAGAAAACTCTAGAAAATGGGGTTCCTGTACTATCTGAGCTTTTAGGAAAA 7227

Qy    164 LysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThrVal 183
          :::   |||      :::|||:::            |||                |||:::
Db   7228 GAAGACACAAGAACAGAGGATCAAGAATTAATGTCTCCTAAAGGT---------ACACTA 7278

Qy    184 LysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluPro 203
          |||:::               |||   |||:::                  :::|||
Db   7279 AAGAGATTTTCATCTCTAGGAAAGGAAAGTCAAGAAGTAGTGAGGCCTTCAAAAGAGGGG 7338

Qy    204 SerValGluThr------------------------------AlaValGlyGluSerAla 213
          ::::::|||:::                                    :::|||
Db   7339 AACCTAGAATCATGGACAGCTTTTAAAGAGGAGAGCCAACACCCACTGGGATTTCCAGGA 7398

Qy    214 SerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeu--------- 230
          :::::::::      |||::::::      ||||||:::::      :::
Db   7399 GCTGAGGACCAGATGCTTGAGAGACTGGTAGAGAAAGAGGATCAGAGCTTCCCAAGGTCT 7458

Qy    231 ---LysGlnGluGlnSerSerThrGluIleProLeuGlnAlaGluSerAspGln---Ala 248
             ::::::|||                  |||||||||   |||:::      :::
Db   7459 CCAGAGGAAGAGGACCAGGAGGCATGTAGACCTCTGCAGAAAGAGAATCAGGAACCACTA 7518

Qy    249 AlaGluGluGluAlaLysAspGluGlyGlu------------------GluLysGlnGlu 262
                |||||||||         ||||||:::                  ||||||:::
Db   7519 GGGTATGAAGAAGCA------GAGGGCCAGATACTTGAGAGACTGATAGAAAAAGAGAGT 7572

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
          :::|||         :::|||||||||                  :::   |||
Db   7573 CAGGAGTCCCTGAGGTCTCCAGAAGAAGAG------------GACCAGGAGGCAGGTAGA 7620
```

```
Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
           |||  :::|||                         ::::::          ::::::
Db   7621 TCTCTGCAGAAA--------------------GAGAATCAGGAGCCACTAGGGTATGAA 7659

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
           :::::::::::         ||||||         :::::::::  ||||||:::   |||
Db   7660 GAAGCAGAGGACCAGATGCTTGAGAGACTGATAGAAAAAGAGAGTCAGGAGTCCCTGAAG 7719

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGlu--------- 339
                 ||||||   :::          :::|||          |||   :::
Db   7720 TCTCCAGAAGAAAACCAGAGGATTGGGAAGCCTCTAGAAAGAGAGAATCAGAAATCTCTG 7779

Qy    340 ---------------------------------ProAlaGluAspThrAspGlnAlaArg 348
                                           |||  |||     :::|||
Db   7780 AGGTATCTTGAAGAAAACCAGGAGACTTTTGTACCACTAGAAAGCAGGAACCAGAGGCCA 7839

Qy    349 LeuSerAla---------------AspTyrGluLysValGluLeuProLeuGlu------ 361
           |||   :::                :::  :::::::::         |||||||||
Db   7840 CTGAGATCTCTAGAAGTAGAAGAGGAGGAGCAGAGAATTGTGAAACCTCTAGAAAAAGTG 7899

Qy    362 ------AspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAla 379
                |||   :::|||   |||           :::||||||        ||||||
Db   7900 AGTCAGGATTCCCTCGGATCTCTA---------GCAGAAGAGAATGTGCAGCCACTGAGG 7950

Qy    380 ThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisVal 399
                    :::   |||:::   :::            :::   :::   |||
Db   7951 ---------TATCTGGAAGAAGATGACTGCATAAATAAGAGCCTTCTAGAA--------- 7992

Qy    400 SerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGlyGluAlaGluGlyGlyVal 419
                    :::||||||   :::      |||         :::      |||
Db   7993 ---------GACAAGACTCACAAGTCCTTGGGGTCTCTTGAAGATAGAAATGGGGATAGC 8043

Qy    420 ValVal-----------GluGlyThrGlyGluSerLeu---ProProGluLys------ 432
           ::::::                    |||      ||||||   |||||||||:::
Db   8044 ATTATTATACCACAAGAAAGTGAGACCCAGGTTTCATTGAGGCCTCCAGAAGAGGAGGAC 8103

Qy    433 --------------------LeuAlaGluProGlnGluValProGlnGluAlaGluPro 445
                                   |||    ||||||        :::      :::|||
Db   8104 CAGAGGATTGTGAACCATCTAGAAAAAGAAAGTCAGGAGTTCTCGAGGTCTTCAGAAGAA 8163

Qy    446 AlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeu 465
           |||::::::|||             |||      :::   |||   :::|||      |||
Db   8164 GAAGAGCAGGTGATG---------GAGAGATCTCTAGAAGGAGAGAACCATGAATCACTG 8214

Qy    466 ThrAspLeuSerProGlu------------GluLysThrLeuProLysHisProGluGly 481
           :::   :::      |||            |||       |||   |||      :::
Db   8215 AGTTCTGTAGAAAAAGAGGACCAGATGGTTGAGAGCCAACTAGAGAAAGAGAGTCAGGAC 8274

Qy    482 IleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySerPro 501
                            :::|||         ||||||||               |||
Db   8275 TCAGGGAAGTCTCTTGAAGAT---GAGAGCCAGGAGACCTTT------------GGACCT 8319

Qy    502 LeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGly 521
           |||:::|||            ::::::         |||:::|||:::   |||
Db   8320 CTGGAAAAA---GAGAATGCAGAGTCCCTGAGATCTCTAGCAGGACAGGACCAA------ 8370

Qy    522 LysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSer 541
                                            |||   |||   :::      |||:::
Db   8371 ---------------------------------GAGGAACAGAAGCTTGAACAAGAGACC 8397

Qy    542 ProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGlu------ 559
             ::::::                      |||      |||   ||||||   |||
Db   8398 CAACAAACACTGAGGGCTGTAGGGAATGAGCAGATGGCAGTGAGCCCACCAGAAAAGGTG 8457

Qy    560 ---ProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGlu 578
              ||||||             |||   |||   ||||||              |||
Db   8458 GATCCAGAG------------TTACCGAAG---CCTCTT---------------GGAAAT 8487

Qy    579 AlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAla 598
             :::|||            |||   |||::::::      :::|||   :::
Db   8488 GACCAGGAAATAGCTAGATCTCTTGGAAAAGAGAATCAAGAGTCACTAGTG--------- 8538

Qy    599 SerPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLys 618
                                                          |||   |||
Db   8539 -------------------------------------------------TCACTGAAA 8547

Qy    619 GluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSer 638
           |||:::   :::|||   ||||||||                              :::
Db   8548 GAAAAAGGTATAGAGACAGTGAAGTCTTTAGAA-----------------------ACA 8583

Qy    639 GluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArg 658
           |||:::   :::   ::::::|||      |||:::            |||   ::::::
Db   8584 GAGATCATAGAACCACTGGAGACTGCAGAAGAGGACCTG---------GAAAGAAGGAAG 8634

Qy    659 ArgValAspThrSerValSer---TrpGluAlaLeuIleCysValGlySerSerLysLys 677
              :::||||||         |||                      |||   :::
Db   8635 TCTATAGATACTCAGGAGCCATTGTGG----------------------TCTACTGAA 8670

Qy    678 ArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGly--- 696
              ||||||:::                  ::::::      |||   :::||||||
Db   8671 GTGGCTAGAGAGACAGTAGAACCTCCAGAAGATGAGCCCCCAGGATCGCTAGGGTCTGTG 8730

Qy    697 AspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValPro 716
           |||   :::|||            |||   :::||||||:::          :::
Db   8731 GATGAGAACCGAGAGACACTGACATCCCTTGAAAAGGAGAGTCA-AGAACTGAGCTCTCT 8789

Qy    717 AlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySer 736
              :::   :::||||||   :::|||   |||:::      ||||||
Db   8790 GGGCAAGTGGAACGTAGAGACCAGGGTAGAGGACAGTCAGCA---GTGCCTGCAAGTAGA 8846

Qy    737 ProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu------ValThr 754
                       |||   ||||||   ||||||   :::   :::
Db   8847 AGAGGGTCTGCAGGAGGAAC-AGCACCAA-GAGTCTCTGAGAGAGGTGAAGCAGGAGCTG 8904

Qy    755 ProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGln 774
           |||         ::::::   :::   |||:::      |||      :::|||   |||
Db   8905 CCTAGCTCTGGAAATCAACAGCGGTGGGAGGATGTGGTGGAGGGCAAAGCAGTGGGTCAG 8964

Qy    775 ---------LeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLys 791
                    |||:::|||            :::   :::::::::   :::   :::
Db   8965 GAAGCACCTCTGGCAACCACAGGAGTGGGAACTGAGGATAAGGCAGAGTTGCATCTG--- 9021

Qy    792 LysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrVal 811
                                         |||   |||   :::|||:::|||
Db   9022 ----------------------AGGGGCAAGGTGGAGAGGAAGAAGCTGCAGCA 9054

Qy    812 GluAspSerGlyProValGluIleAsnGluAsp-------------------------- 822
           |||            |||:::   :::|||
Db   9055 GAGGGA------------GAGCTGTTGCAGGATATTGTGGGGAGGCCTGGAGTCTGGGG 9102

Qy    823 ---------------AspProAsnValProAla------------------------- 828
                          :::      ||||||||
Db   9103 AGCTCTGAGCCCAAGGAGCAGAGGGTCCCTGCTGAGGCCCTCGACAACCTGGAAGGAGGG 9162

Qy    829 ------ValValProLeuSerGluTyr------AsnAlaValGluArgGluLysMetGlu 844
                 ||||||:::::::::                ||||||::::::
Db   9163 GCCTTAGAGGTCCCAGTTGCTCAGTCAATGCCAGAGGTGACAGAGCGAGATGAGGATAGA 9222

Qy    845 AlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeu 864
           ||||||                                          :::   :::
Db   9223 GCCCAAGCAGGTGAACAA-----------------------------------GACTCCATA 9249

Qy    865 SerLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThr 884
              ||||||                    ::::::      |||
Db   9250 GAGGTGACCCTT--------------------GGGTTAGAGGCTGCCAGAACTGGACTG 9288

Qy    885 SerValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHis 904
              :::|||:::                           :::   :::|||
Db   9289 GAACTCGAGCAGGAA---------------GTGGTAGGGCTAGAGGACCCAAGGCATTTT 9333

Qy    905 ThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGlu 924
                 ||||||:::||||||            :::   |||:::   :::   |||:::
Db   9334 GCCAGGGAGGAGGCCATTCCCCCA------TCCCTGGGGGAGGAAAGTGTGAAGGCAAAG 9387

Qy    925 GluThrProValLeuThrGlnThrLeu-----------------ProGluGlyLysAsp 938
```

```
Db     9388 ------------ATAGCTCAGGGCTTGGAAGGGCCTGGAAAGGAACCAAAAGAGGCAGGT 9435

Qy      939 AlaHisAspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThr 958
Db     9436 GCTCTGGACTCGGGGATCCTTGAATTGCCCAAGACTAGCAGCGAGGCTCTGGAATGCCAG 9495

Qy      959 ------GluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGlu 976
Db     9496 GGCCATGAAGAGTCTGAGTCCATGGAGGGCTGGGAAGAAGAGGAGGCCTCA------CTG 9549

Qy      977 GluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGlu 996
Db     9550 GAGACTTCAGAT---------------CATGAGGGCAGTGATGCCCCTCAGCCCAGGCCC 9594

Qy      997 GluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArg 1016
Db     9595 CCAGAAACA------------------------------------GAAGAAGATGAG 9615

Qy     1017 GlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGlu--------------- 1031
Db     9616 GGTGCACAGGCAGCACTGACAGCCCCTGGTCCCAAGCTCTTGGAACCCTGTTCACCCATC 9675

Qy     1032 ---------------------GluSerGlnValProAlaThrGlnThrValGln------ 1042
Db     9676 CCAATCCTGACAGATGCCCATGAGCTGCAG---CCCCAGGCTGAGGGGATCCAGGAGGCT 9732

Qy     1043 ------------ArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAsp 1058
Db     9733 GGCTGGCAGCCAGAAGCTGGGTCTGAAGCACTAGAAAGGGTAGAAAATGAGCCAGAG--- 9789

Qy     1059 SerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGlu 1078
Db     9790 ------------------------------------TTTGGTCTTGGGGAGATCCCGGAG 9813

Qy     1079 AlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeu 1098
Db     9814 ---GGCCTCCAGGATTGGGAAGAGGGCAGAGAAGAAAGCGAGGCAGATGATCTAGGGGAA 9870

Qy     1099 GluValProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeu 1118
Db     9871 ACTCTCCCTGACTCTACTCCCCTGGGCCTCTACCTGAGGTCCCCTGCTTCTCCAAAGTGG 9930

Qy     1119 GlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGlu 1138
Db     9931 GATCTGGCTGGAGAACAGAGGCTTTCCCCTCAAGGGGATGCCGGGAAGGAAGAC------ 9984

Qy     1139 ThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGlu 1158
Db     9985 TGGGGTCCTGCTGTCCCCGCTGCCCAGGGCCTCAGTGGTCCACCGGAAGAGGAGGAGGAG 10044

Qy     1159 ThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGlu 1178
Db    10045 CAAGGCCATGGCTCTGAC------------------------CTATCATCTGAGGAGTTT 10080

Qy     1179 GluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAla 1198
Db    10081 GAGGACCTAGGGACT------------GAGGCCTCTCTTCTTCCA---GGGGTTCCCAAG 10125

Qy     1199 Gln------GluGluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGln 1216
Db    10126 GAGGTGGCAGATCACGTGGGCCAAGTGCCC------CCGGTACTGCAGCCTGCATGCTGG 10179

Qy     1217 GluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGlu 1236
Db    10180 GATCAGGGTGGGGAATCTGATGGGTTTGCTGATGAGGAAGAAAGTGGGGAGGAGGGAGAG 10239

Qy     1237 ValAspTrpLeuAspGlyGluLysValLysGluGluGlnGluValPheValHisSerGly 1256
Db    10240 GAAGAAGATGCTGAT---GAGGAAGGAGCAGAGTCAGGAGCTCAGTGGTGGGGTCAGGG 10296

Qy     1257 ProAsnSerGln------LysAlaAlaAspValThrTyrAsp------------SerGlu 1270
Db    10297 GCCTCTGGTGGAGGCTGCAAGGTCCAGGATATTGCCCAAAGAGGAGACCCGGTACAGGAG 10356

Qy     1271 ValMetGlyValAlaGlyCysGlnGluLys------------GluSerThrGluValGln 1286
Db    10357 TCTGTGGGTGTCAGTGGTCTCTGGGATGATGGCTTGAGAGGTGCTGCAGCTAATGTTCCT 10416

Qy     1287 SerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThr 1306
Db    10417 GCCCTAGAGATG------------------------------------------GTA 10431

Qy     1307 LysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGly 1326
Db    10432 TCTCAGGACAGTGCTGAGCCTTCTGGGTCAGAGGAGTCTGAGTCTGCTTCCTTGGAGGGG 10491

Qy     1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArg------------ 1342
Db    10492 GAGGAAGGTCAAGTGACTGACCATTTAGATGCTCCCCAGGAGGTGACCAGCATGGTCCCG 10551

Qy     1343 ------GlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
Db    10552 GGGGTAGGAGATGCCTTTGACATTGGTGGCCAGAGCCCCAACTTGGACTCAGAACAAGTG 10611

Qy     1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
Db    10612 AATGGGAAAATGGAGAATGGACTAGAACAGGCTGAGGGGCAGGTGGTCCTGGATGGGGAC 10671

Qy     1381 AlaValGluLysValIleGluThrValValIleSerGluThrGly--------------- 1395
Db    10672 GAGGATCAAGAACTCCTA------TTACAGGGACAGGAGGTGGTGCTCTAAAGGTTCCT 10725

Qy     1396 ------GluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys--------- 1410
Db    10726 TTGGTAGCATCTCCT------------GTGCATCTAGGCCCAAGCCAGCCCCTGAAGTTC 10773

Qy     1411 ------SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValPro 1428
Db    10774 ACTCTGAGTGGGGTAGATGGGGATTCCTGGTC-CTCAGGGGAAGACTAGAAACTGCCCCT 10832

Qy     1429 Leu 1429
Db    10833 CTG 10835
```

RESULT 15
US-08-056-200-93
; Sequence 93, Application US/08056200
; Patent No. 5616500
; GENERAL INFORMATION:
; APPLICANT: Steinert, Peter M.
; APPLICANT: Lee, Seung-Chul
; APPLICANT: Kim, In-Gyu
; APPLICANT: Chung, Soo-Il
; APPLICANT: Park, Sang-Chul
; TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and
; TITLE OF INVENTION: Mehods of Using Same
; NUMBER OF SEQUENCES: 117
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, Sixteenth Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: U.S.A.
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/056,200
; FILING DATE: 30-APR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fedrick, Michael F.

```
;      REGISTRATION NUMBER:  36,799
;      REFERENCE/DOCKET NUMBER:  NIH054.001A
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (714) 760-0404
;      TELEFAX:  (714) 760-9502
;  INFORMATION FOR SEQ ID NO:  93:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  9551 base pairs
;      TYPE:  nucleic acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  cDNA
;    HYPOTHETICAL:  NO
;    ANTI-SENSE:  NO
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  1507..1644
;    FEATURE:
;      NAME/KEY:  intron
;      LOCATION:  1645..2511
;    FEATURE:
;      NAME/KEY:  CDS
;      LOCATION:  2512..8070
US-08-056-200-93

Alignment Scores:
Pred. No.:               8.86e-16      Length:         9551
Score:                   366.00        Matches:        347
Percent Similarity:      34.29%        Conservative:   251
Best Local Similarity:   19.90%        Mismatches:     701
Query Match:             4.53%         Indels:         452
DB:                      1             Gaps:           60

US-09-902-432-4 (1-1596) x US-08-056-200-93 (1-9551)

Qy     74 GluGluAsnGlnGluGlyGlnGlu------GluGluValValAspGluAspValGlyGln 91
          ||||||:::|||          |||      :::     ::::::|||:::   |||||
Db   2719 GAAGAAGACCAAAGGAGATTCGAGCCCCGGGACAGACAACTGGAAGAAGAACCTGGGCAA 2778

Qy     92 ArgGluSerGluAspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThr 111
          |||      :::      :::|||::::::|||      |||:::|||
Db   2779 CGACGCAGGCAGAAGAGGCAGGAACAGGAGAGG------GAGCTAGCTGAGGGA------ 2826

Qy    112 AlaValGluAspIleThrLysAspGlyGlnGluGluThrSerGluIleIleGluGlnIle 131
                                  :::   |||   |||      |||   :::||||||
Db   2827 ---------------------GAGGAGCAAAGTGAGAAACAAGAGCGACTTGAACAGCGC 2865

Qy    132 ProAlaSerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAsp 151
                :::   |||            |||      |||   |||            :::
Db   2866 GACAGGCAGCGCCGCGACGAGGAGCTGTGGCGGCAAAGGCAAGAATGGCAAGAACGGGAA 2925

Qy    152 ValGlyPheLysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLys 171

Db   2925 ------------------------------------------------------------ 2925

Qy    172 AsnGluLysSerAspThrValGlnLeuLeuThrValLys--------------------- 184
                :::::::::      ||||||   :::   |||
Db   2926 GAGCGCCGTGCAGAGGAAGAGCAGCTGCAGAGTTGCAAAGGTCACGAAACTGAGGAGTTT 2985

Qy    185 LysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSer 204
             ||||||   :::   |||:::|||                  :::   :::   |||
Db   2986 CCAGACGAAGAGCAACT-GCGAAGGCGGGAGCTGCTGGAGCTGAGGAGGAAGGGCCGCGA 3044

Qy    205 ValGluThrAlaValGlyGluSerAlaSerLysGluSer--------------------- 217
          |||||||||         |||   ::::::|||
Db   3045 GGAGAAACAGCAGCAAAGGCGAGAGCGCCAAGACAGAGTGTTCCAGGAGGAAGAAGAGAA 3104

Qy    218 ---------GluLeuLysGlnSerThr-GluLysGlnGluGlyThrLeuLysGlnGluGl 234
                   |||   :::   |||      |||:::|||      |||::::::|||::
Db   3105 AGAGTGGAGGAAGCGCGAGACAGTGCTCCGGAAGGAAGAAGAGAAGTTGCAGGAAGAGGA 3164

Qy    234 nSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLy 254
           :        :::         ||||||   |||   :::|||         :::   |||   ::
Db   3165 GCCGCAGCGGCAAAGAGAGCTCCAGGAGGAAGAAGAGCAGCTACGGAAGCTGGAGCGGCA 3224

Qy    254 sAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerProSerSe 274
           ::::            |||:::||||||:::|||         :::
Db   3225 AGAGCTGAGGAGGGAGCGCCAGGAGGAAGAGCAGCAGCAGCAAAGGCTGAGGCGCGAGCA 3284

Qy    274 rProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGl 294
             :::
Db   3285 GCAACTAAGG-------------------------------------------------- 3294

Qy    294 yTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLy 314
               ||||||:::            ::::::   ::::::::::::         :::::
Db   3295 ----CGCAAGCAGGAGGAGGAGAGGCGCGAGCAGCAGGAGGAGAGGCGCGAGCAGCAGGA 3350

Qy    314 sArgLysGluGlnGluAlaGluLysValAsp--------------GluGluGluLysGl 329
           :|||:::||||||:::   |||:::   :::               |||:::|||:::
Db   3351 GAGGCGCGAGCAGCAGGAGGAGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGGAGGAGAG 3410

Qy    329 uLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArgLe 349
            :::   :::         |||||||||            |||   |||
Db   3411 GCGCGAGCAGCAGCTGAGGCGCGAGCAGGAGGAGGAGAGGCGCGAGCAGCAGCTGAGGCG 3470

Qy    349 uSerAlaAspTyrGluLysValGluLeuProLeu------GluAspGlnValGlyAspLe 367
              :::   |||:::   |||      |||      |||   :::   |||:::
Db   3471 CGAGCAGGAGGAGGAGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGGAGGA-GGAGAGGC 3529

Qy    367 uGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMe 387
              |||::::::|||   :::   |||      |||            :::   :::
Db   3530 GCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGC 3589

Qy    387 tGluAlaHisGlnGluValValAlaGluValHisValSerThrValGluLys-------- 404
             |||         ||||||         :::         :::
Db   3590 AGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGC 3649

Qy    405 ---------------------------------------------------ThrGluGl 407
                                                              |||
Db   3650 TGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGCGCGAGCAGCAGCTGAGGC 3709

Qy    407 uGluGlnGlyGlyGly-----------------------GlyGluAlaGluGlyGlyVa 419
                |||||||||                       |||||||||
Db   3710 GCGAGCAGGAGGAGGAGAGGCACGAGCAGAAGCACGAGCAGGAGAGGCGCGAGCAGCGGC 3769

Qy    419 lValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluVa 439
                          ||||||:::            :::|||||||||   :::
Db   3770 TGAAGCGCGAGCAGGAGGAGAGGCG------------CGATTGGCTGAAGCGCGAGGAGG 3817

Qy    439 lProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGl 459
                ||||||   |||   |||         :::   |||         ||
Db   3818 AGACGGAGAGGCACGA---GCAGGAGAGGCGCAAGCAGCAGCTGAAGCGCGACCAGGAGG 3874

Qy    459 yGlyAspHisThrGlnLeuThrAsp-LeuSerProGluGluLys---------------- 473
           |||:::         |||   :::   |||      ||||||:::
Db   3875 AGGAGAGGCGCGAACGTTGGCTGAAGCTCGAGGAGGAGGAGAGGCGCGAGCAGCAGGAGA 3934

Qy    474 -----------------------------------------------ThrLeuProLysH 478
                                                         |||   :::
Db   3935 GGCGCGAGCAGCAACTAAGGCGGGAGCAAGAGGAGAGGCGCGAGCAGCGGCTGAAGCGCC 3994

Qy    478 isProGluGlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValG 498
              |||               :::   :::   |||   |||:::      :::   |
Db   3995 AGGAGGAGGAAGAGAGGCTCCAGCAGCGGTTGAGGAGCGAGCAACAACTAAGACGCGAGC 4054

Qy    498 lnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeu---------- 514
            ||         |||::::::|||                  |||:::|||
Db   4055 AGGAGGAGAGGCTCGAGCAGCTGCTGAAGCGCGAGGAGGAGAAGAGGCTCGAGCAGGAGA 4114

Qy    514 ------------------------------------------------------------ 514

Db   4115 GGCGAGAGCAGCGGCTGAAGCGCGAGCAGGAGGAGAGGCGCGATCAGCTGCTGAAGCGCG 4174

Qy    515 -----SerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGlyG 533
                      ::::::|||:::   ||||||            ::::::|||
Db   4175 AGGAGGAGAGGCGCCAGCAGCGGCTGAAGCGC-----------GAGCAGGAAGAGAGGC 4222
```

```
Qy    533 luTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSerS 553
Db   4223 TCGAGCAGCGACTGAAGCGCGAGGAGGTGGAGAGACTCGAGCAGGAGGAGAGGCGCGACG 4282

Qy    553 erAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluA 573
Db   4283 AGCGGCTGAAGCGCGAGGAGCCGGAGGAAGAGAGGCGCCACGAGCTGCTGAAGAGCGAGG 4342

Qy    573 laProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluG 593
Db   4343 AGCAGGAGGAGAGGCGCCACGAGCAACTGAGGCGCGAGCAGCAGGAAAGGCGCGAGCAGC 4402

Qy    593 lyIleThrProTrpAlaSer--------------------PheLysLysMetValThrP 606
Db   4403 GGCTGAAGCGCGAGGAGGAGGAAGAGAGGCTCGAGCAGCGGCTGAAGCGCGAGCATGAGG 4462

Qy    606 roLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValL 626
Db   4463 AAGAGAGGCGCGAGCAGGAGCTAGCTGAGGAGGAGCAGGAACAGGCCCGGGAGCGGATTA 4522

Qy    626 ysSer--------------------AlaThrLeuSerSerThrAspSerThrValSerG 639
Db   4523 AGAGCCGCATCCCGAAGTGGCAGTGGCAGCTAGAAAGCGAGGCCGACGCACGGCAAAGCA 4582

Qy    639 luMetGlnAspGluValLysThrValGlyGluGluGlnLysPro------GluGluProL 657
Db   4583 AAGTCTTACTCGAGGCCCCGCAAGCAGGAAGGGCAGAGGCGCCGCAAGAGCAGGAGGAAA 4642

Qy    657 ysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysL 677
Db   4643 AGAGGCGGCGCGAGAGTGAGCTGCAATGGCAG--------------------GAGGAGG 4681

Qy    677 ysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyA 697
Db   4682 AACGGGCTCACCGGCAGCAGCAGGAAGAGGAGCAGCGCCGGGACTTCACATGG------- 4734

Qy    697 spSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProA 717
Db   4735 --CAGTGGCAGGCGGAGGAAAAGAGC---------GAGAGGGGCCGTCAGAGGCTGTCGG 4783

Qy    717 laSerThrGlnGluGlnAspGln-AlaGlnGlySerSerSerProGluProAlaGlySer 736
Db   4784 CCAGGCCCCCATTGCGGGAGCAGCGGGAGAGGCAGCTGAGGGCCGAGGAGCGCCAGCAGC 4843

Qy    737 ProSerGluGly---------------------------------------------- 740
Db   4844 GGGAACAACGGTTTCTCCCGGAGGAGGAGGAGAAGGAGCAGCGCGGCCGCCAGCGACGCG 4903

Qy    741 ---------GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLys 757
Db   4904 AGAGGGAGAAAGAGCTGCAGTTCCTGGAGG------AAGAGGAGCAGCTCCAG-CGGCGG 4956

Qy    758 LysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThr 777
Db   4957 GAGCGTGCCCAACAGCTCCAGGAGGAGGAGGACGGCCTCCAGGAGGATCAGGAGAGGAGG 5016

Qy    778 GluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArg 797
Db   5017 CGACAGGAGCAGCGCCGCGACCAAAAATGG---------AGGTGGCAACTAGAAGAAGAA 5067

Qy    798 ArgLysLysArg-----------------AlaAspGlyLysGlnGluGlnAlaThrVal 811
Db   5068 AGGAAGAGACGCCGCCACACGCTGTACGCCAAGCCAGCCCTACAAGAGCAGCTGAGGAAG 5127

Qy    812 GluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValValPro 831
Db   5128 GAACAGCAG---------CTGCTGCAGGAGGAGGAG----------------------- 5154

Qy    832 LeuSerGluTyrAsnAlaValGluArgGluLysMet----------------------- 843
Db   5155 ---GAGGAGCTACAGAGAGAGGAGCGCGAGAAGAGAAGGCGCCAAGAACAGGAGAGACAA 5211

Qy    844 ---------GluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
Db   5212 TACCGCGAGGAAGAGCAGCTGCAGCAGGAGGAAGAGCAGCTGCTG--------------- 5256

Qy    861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
Db   5257 AGAGAGGAACGGGAGAAAAGAAG-ACGCCAGGAGCGGGAAAGGCAATATCGGAAGGATAA 5315

Qy    881 ArgAlaValThr--------------------------SerValGluGluArgSerPro 891
Db   5316 GAAGCTGCAGCAGAAGGAAGAGCAGCTGCTGGGAGAGGAACCGGAGAAGAGAAGGCGCCA 5375

Qy    892 SerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetPro 911
Db   5376 GGA------------------------------------GCGGGAGAAAAAATACCG 5396

Qy    912 ProValGluGluVal--------------------ThrGluLysAspIleIleAlaGlu 924
Db   5397 CGAGGAAGAGGAGTTGCAGCAGGAGGAAGAGCAGCTGCTGAGAGAGGAACGGGAGAAGAG 5456

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   5457 AAGGCGCCAGGAGTGGGAGAGGCAGTACCGCAAAAAAGACGAGCT--------------- 5501

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThr-GluThrSerGluAlaLe 964
Db   5502 -----------------------------------GCAGCAGGAAGAAGAGCAGCTGCT 5525

Qy    964 uArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAl 984
Db   5526 GAGAGAGGAACGGGAGAAAAGA-------------------------------------- 5547

Qy    984 aValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVa 1004
Db   5548 -------AGACTCCAGGAGCGGGAGAGGCAATATCGGGAGGAAGAGGAGCTGCAGCAGGA 5600

Qy   1004 lGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIle-------- 1021
Db   5601 GGAAGAGCAGCTGCTGGGAGAGGAACGGGAGACGAGAAGGCGCCAGGAGCTGGAGAGGCA 5660

Qy   1022 -LeuGlnAlaValAlaAspLysValLysGluGluSerGlnVal-----------ProAl 1037
Db   5661 ATATCGGAAGGAAGAGGAGCTGCAGCAGGAGGAAGAGCAGCTGCTGAGAGAGGAACCGGA 5720

Qy   1037 aThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGl 1057
Db   5721 GAAGAGAAGGCGCCAGGAGCGGGAGAGGCAATGTCGGGAGGAAGAGGAGCTGCAGCAGGA 5780

Qy   1057 uAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGl 1077
Db   5781 GGAAGAGCAGCTGCTGAGAGAGGAACGGGAGAAGAG-AAGGCGCCA-------------- 5825

Qy   1077 nGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSe 1097
Db   5826 -------GGAGCTGGAGAGGCAATATCGGGAGGAGGAAGAGCTTCAGCGCCAGAAAAGGA 5878

Qy   1097 rLeuGluValProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLy 1117
Db   5879 AACAGCGATACCGGGATGAGGATCA---------------------------------- 5903

Qy   1117 sLeuGlnGlnLeuMetGluGlnAlaValAla-ProGluSerSerGluThrLeuThrAspS 1137
Db   5904 ----GCGCAGTGATCTGAAATGGCAGTGGGAACCAGAAAAAGAAAATGCAGTTCGTGATA 5959

Qy   1137 er--------------------GluThrAsnGlySerThrProLeuAlaAspSerAspT 1150
Db   5960 ACAAGGTTTACTGCAAAGGCAGAGAGAATGAACAGTTCCGGCAGTTGGAAGATTCCCAGG 6019

Qy   1150 hrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaA 1170
Db   6020 TGCGCGACAGACAATCCCAG------------------CAAGATCTGCAGCACCTGCTGG 6061

Qy   1170 laValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProS 1190
```

```
                 :::|||   :::       |||:::|||            |||:::
Db   6062 GTGAACAGCAAGAGAGAGATCGTGAGCAAGAGAGGAGGCGCTGGCAGCAGGCCAACAGGC 6121

Qy   1190 erThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspV 1210
                                   :::||||||      |||       ||||||
Db   6122 ATTTCCCAGAGGAAGAACAGCTGGAGCGAGAAGAGCAAAAGGAAGCCAAAAGGCGCGACA 6181

Qy   1210 alLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrG 1230
                           :::::::::|||                     :::    |
Db   6182 GGAAGTCCCAAGAGGAAAAGCAGTTGCTGAGAGAG--------------GAAAGAGAAG 6226

Qy   1230 luValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGluGluGlnG 1250
           ||    :::      |||   |||               |||   :::||||||:::
Db   6227 AGAAGAGACGCCGTCAAGAGACAGAC------------AGAAAATTCCGCGAGGAGGAAC 6274

Qy   1250 luValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerG 1270
          :::::                        ::::::                        |
Db   6275 AGCTGCTC------------------CAGGAAAGGGAGGAACAGCCGCTGCTCCGCCAAG 6316

Qy   1270 luValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSerLeuSerL 1290
           ||                     ::::::::::::      |||   :::   |||
Db   6317 AG--------------------CGTGACAGAAAATTCCGCGAAGAGGAACTGCTCCATC 6355

Qy   1290 euGluGluGly-----------------GluMetGluThrAspValGluLysGluLysA 1304
             |||:::|||                 |||                |||:::|||:::
Db   6356 AGGAACAAGGGAGAAAATTCCTCGAGGAGGAACAGCGGCTGCGCGAGGAACGGGAGAGAA 6415

Qy   1304 rgGluThrLysProGluGlnVal------SerGluGluGlyGluGlnGluThrAlaAlaP 1322
           ::      |||  ||||||           ||||||  ||||||       :::   |
Db   6416 AATTCCTTAAGGAGGAACAGCAGCTGCGCCTCGAGGAGCGCGAGCAACT-GCGTCAGGAC 6474

Qy   1322 roGluHisGlu-----------GlyThrTyrGlyLysProValLeu------------ 1333
           ||       :::             ||||||      :::|||
Db   6475 CGCGACAGAAAATTCCGCGAGGAGGAACAGCAGCTGAGCCGCCAAGAGCGTGACAGAAAA 6534

Qy   1334 -------------ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerL 1349
                          |||       |||       |||   |||    |||
Db   6535 TTCCGTGAAGAGGAACAGCAGGTGCGCCGCCAGGAACGAGAGAGAAAATTCCTGGA---- 6590

Qy   1349 euGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValG 1369
             ||||||:::   :::   |||          :::                ::::::
Db   6591 --GGAGGAACAGCAGCTGCGCCAGGAGCGTCACAG-----------------AAAATTC 6630

Qy   1369 lnSerLeuAspThrThrValThr---------GlnThrAlaGluAlaValGluLysValI 1386
                   |||                    :::||||||   |||      :::
Db   6631 CGCGAAGAGGAACAGCTGCTCCAGGAAAGGGAAGAACAGCAGCTGCACCGCCAAGAGCGT 6690

Qy   1386 leGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuL 1406
             :::   :::            |||      :::|||
Db   6691 GACAGAAAATTCCTGGAGGAGGAACAACAGCTGCGCCG--------------------- 6728

Qy   1406 euProAla------GluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHis---A 1423
             |||          :::|||         ||||||         |||   |||
Db   6729 --CCAAGAGCGTGACAGAAAATTCCGCGAACAGGA---------ACTGCGCAGTCAGGAA 6777

Qy   1423 laGluAspThrValProLeuGly--------------ProGluSerGlnAlaGluSerI 1438
               :::   :::|||   |||                 |||   :::   ||||||   |
Db   6778 CCAGAGAGAAAATTCCTCGAGGAGGAACAGCAGCTGCACCGCCAGCAACGGCAGAGAAAA 6837

Qy   1438 lePro-----------IleIleValThrProAlaProGluSerThrLeu-HisProAsp 1453
           |||||                        |||          :::|||         |||
Db   6838 TTCCTCCAGGAGGAACAGCAGCTGCGCCGCCAGGAGCGCGGGCAACAGCGGCGTCAGGAC 6897

Qy   1454 LeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysProAsp 1473
                :::               ::::::   |||   |||   |||::::::
Db   6898 CGTGACAGAAAATTCCGCGAGGAGGAACAGCTGCGCCAGGAGAGGGAGGAACAGCAGCTG 6957

Qy   1474 AlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeuLysAlaGlu 1493
            :::      :::   |||   |||            :::||||||   :::   |||
Db   6958 AGCCGCCAAGAGCGTGACAGAAAATTCCGTTTAGAGGAACAGAAAGTGCGCCGCCAGGAA 7017

Qy   1494 ProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnValIleGlnThr 1513
             |||          :::|||   :::
Db   7018 CAAGAGAGAAAATTCATGGAGGACGAA-------------------------------- 7044

Qy   1514 AlaValAspGlnPheAlaArgThrGluThrAlaProGluThrHisAlaTyrAspSerGln 1533
                   |||      |||   |||            :::         |||   :::
Db   7045 ------CAGCAGCTGCGCCGCCAGGAGGGCCAACAACAGCTGCGCCAGGAGGACAGAAAA 7098

Qy   1534 ThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCysTrpThrLysMet 1553
             :::                   |||      ||||||      :::       :::
Db   7099 TTCCGCGAAGACGAACAGCTGCTCCAGGAAAGGGAAGAACAGCAGCTGCACCGCCAAGAG 7158

Qy   1554 LysAspAlaLysMetLysHisProValProGlnProArgGluAspLeuGlnValLeuThr 1573
          :::|||   |||                ||||||   |||          :::      |||
Db   7159 CGTGACAGAAAATTCCTCGAGGAGGAACCGCAGCTGCGCCGCCAGGAGCGCGA-ACAACA 7217

Qy   1574 ValLeuGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaPro 1593
                   |||   ::::::||||||:::   :::|||                ||||||
Db   7218 ---------GCTGCGTCACGACCGCGACAGAAAATTCCGTGAAGAGGAACAGCTGCTCCA 7268
```

Search completed: December 13, 2002, 04:17:51
Job time : 438 secs

GenCore version 5.1.3


```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         December 13, 2002, 03:13:21 ; Search time 944 Seconds
                                              (without alignments)
                                              668.597 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       355320 seqs, 197730502 residues

Total number of hits satisfying chosen parameters:      710640

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09902432/runat_12122002_141453_25918/app_query.fasta_1.1735
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09902432_@CGN_1_1_85_@runat_12122002_141453_25918
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      Published_Applications_NA:*
                1:  /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
                2:  /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
                3:  /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
                4:  /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
                5:  /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
                6:  /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
                7:  /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
                8:  /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
                9:  /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
                10:  /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
                11:  /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
                12:  /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
                13:  /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
                14:  /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		8073	100.0	6160	9	US-09-902-432-3	Sequence 3, Appli
2		6005	74.4	5134	9	US-09-902-432-1	Sequence 1, Appli
3		4300.5	53.3	5346	10	US-09-738-877-2	Sequence 2, Appli
4		4300.5	53.3	6608	10	US-09-738-877-1	Sequence 1, Appli
5		4300.5	53.3	6608	10	US-09-880-107-3439	Sequence 3439, Ap
6		462	5.7	5361	9	US-09-742-096-2	Sequence 2, Appli
7		462	5.7	6152	9	US-09-742-096-1	Sequence 1, Appli
8		431.5	5.3	7997	10	US-09-864-761-17726	Sequence 17726, A
9		427	5.3	15231	10	US-09-917-800A-1505	Sequence 1505, Ap
10		418	5.2	14800	10	US-09-954-456-1601	Sequence 1601, Ap
11		411	5.1	7104	10	US-09-815-242-4580	Sequence 4580, Ap
12		411	5.1	7107	10	US-09-815-242-8291	Sequence 8291, Ap
13		408	5.1	10300	10	US-09-960-253-145	Sequence 145, App
14		398.5	4.9	4840	10	US-09-880-107-3423	Sequence 3423, Ap
15		395	4.9	11185	10	US-09-880-107-3311	Sequence 3311, Ap
16	c	388	4.8	3953	10	US-09-864-761-19041	Sequence 19041, A
17		382.5	4.7	7639	9	US-09-971-536-28	Sequence 28, Appl
18		382.5	4.7	7792	12	US-10-044-090-359	Sequence 359, App
19		379	4.7	7497	10	US-09-960-253-175	Sequence 175, App
20		378.5	4.7	10432	10	US-09-919-172-97	Sequence 97, Appl
21		377	4.7	7434	10	US-09-815-242-4761	Sequence 4761, Ap
22		377	4.7	7437	10	US-09-815-242-8869	Sequence 8869, Ap
23		371	4.6	6940	10	US-09-880-107-1724	Sequence 1724, Ap
24		365.5	4.5	5331	10	US-09-764-176-6	Sequence 6, Appli
25		355	4.4	4854	9	US-09-963-875-2	Sequence 2, Appli
26		351.5	4.4	1855	10	US-09-864-761-19708	Sequence 19708, A
27		346.5	4.3	4372	10	US-09-864-864-323	Sequence 323, App
28		346.5	4.3	4372	10	US-09-864-864-337	Sequence 337, App
29		345	4.3	10625	10	US-09-727-384-5	Sequence 5, Appli
30		342.5	4.2	7065	9	US-09-991-496-115	Sequence 115, App
31		342.5	4.2	7065	10	US-09-874-923-115	Sequence 115, App
32		341	4.2	3388	9	US-09-954-531-988	Sequence 988, App
33		341	4.2	3388	9	US-09-954-531-1382	Sequence 1382, Ap
34		341	4.2	3388	10	US-09-954-456-1602	Sequence 1602, Ap
35		341	4.2	3388	10	US-09-967-768A-245	Sequence 245, App
36		335.5	4.2	4780	10	US-09-962-436-287	Sequence 287, App
37		335.5	4.2	7596	10	US-09-954-456-2215	Sequence 2215, Ap
38		335	4.1	1958	10	US-09-864-761-2927	Sequence 2927, Ap
39		334.5	4.1	7035	10	US-09-815-242-8615	Sequence 8615, Ap
40		329.5	4.1	6114	9	US-09-938-842A-2700	Sequence 2700, Ap
41		327	4.1	6457	10	US-09-880-107-3389	Sequence 3389, Ap
42		323.5	4.0	5721	10	US-09-785-770A-15	Sequence 15, Appl
43		323.5	4.0	8121	10	US-09-785-770A-14	Sequence 14, Appl
44		323	4.0	7869	10	US-09-954-456-1921	Sequence 1921, Ap
45		316	3.9	4747	9	US-09-854-133-729	Sequence 729, App

ALIGNMENTS

```
RESULT 1
US-09-902-432-3
; Sequence 3, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 6160
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-902-432-3

Alignment Scores:
Pred. No.:             0              Length:        6160
Score:                 8073.00        Matches:       1596
```

```
Percent Similarity:       100.00%       Conservative:   0
Best Local Similarity:    100.00%       Mismatches:     0
Query Match:              100.00%       Indels:         0
DB:                       9             Gaps:           0

US-09-902-432-4 (1-1596) x US-09-902-432-3 (1-6160)

Qy        1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThr 20
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       33 ATGGGCGCAGGCAGTTCCACCGAGCAGCGGAGCCCCGAGCAGCCGGCGGGGAGCGACACG 92

Qy       21 ProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGly 40
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       93 CCGAGCGAGCTGGTGCTCAGTGGCCATGGCCCGCAGCTGAAGCCTCGGGAGCAGCTGGA 152

Qy       41 AspProAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSer 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      153 GACCCCGCCGACGCGGACCCCGCCACCAAGCTCCCACAGAAGAATGGCCAGCTGTCTTCT 212

Qy       61 ValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      213 GTCAACGGCGTAGCTGAACAAGGAGATGTCCATGTCCAAGAGGAAAACCAGGAGGGGCAG 272

Qy       81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      273 GAGGAAGAAGTCGTTGATGAGGATGTTGGACAGCGAGAGTCAGAAGATGTGAGAGAAAAA 332

Qy      101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      333 GACCGAGTTGAAGAAATGGCGGCCAACTCCACAGCTGTTGAAGATATCACAAAGGATGGG 392

Qy      121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      393 CAGGAGGAGACATCAGAAATAATTGAACAGATCCCTGCTTCAGAAAACAATGTGGAAGAA 452

Qy      141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      453 ATGGTACAGCCTGCTGAGTCCCAGGCTAATGATGTTGGCTTCAAGAAAGTATTTAAATTT 512

Qy      161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      513 GTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAATGAAAAGTCAGATACTGTCCAACTA 572

Qy      181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      573 CTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCTGTCGGAGCTGGAGACCAC 632

Qy      201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      633 CAGGAGCCCAGTGTGGAGACTGCCGTCGGAGAGTCAGCATCCAAAGAAAGTGAGCTGAAG 692

Qy      221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      693 CAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGAACAGAGCAGCACAGAAATCCCC 752

Qy      241 LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLys 260
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      753 CTTCAAGCCGAATCTGATCAAGCGGCTGAGGAAGAAGCCAAAGATGAAGGAGAAGAAAAA 812

Qy      261 GlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr 280
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      813 CAAGAGAAAGAGCCCACCAAGTCCCCAGAATCCCCGAGCAGCCCAGTCAACAGTGAGACA 872

Qy      281 ThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSer 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      873 ACATCTTCCTTCAAGAAGTTCTTCACTCACGGTTGGGCCGGCTGGCGCAAGAAGACCAGC 932

Qy      301 PheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAla 320
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      933 TTCAAGAAATCAAAAGAGGATGATCTGGAAACTGCCGAGAAGAGAAAGGAGCAAGAGGCA 992

Qy      321 GluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro 340
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      993 GAAAAAGTAGACGAGGAAGAAAAGGAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCG 1052

Qy      341 AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeu 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1053 GCAGAAGACACAGACCAGGCCAGGTTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTG 1112

Qy      361 GluAspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThr 380
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1113 GAAGACCAGGTTGGTGACCTGGAGGCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACG 1172

Qy      381 GluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSer 400
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1173 GAAGTGTTTGATGAGAAGATGGAAGCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGC 1232

Qy      401 ThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1233 ACCGTGGAGAAGACAGAGGAGGAGCAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTG 1292

Qy      421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1293 GTAGAAGGAACAGGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCC 1352

Qy      441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1353 CAGGAAGCTGAGCCTGCTGAGGAGCTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGA 1412

Qy      461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1413 GACCACACTCAACTGACAGACCTAAGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAA 1472

Qy      481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1473 GGCATTGTCAGTGAGGTGGAGATGCTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGT 1532

Qy      501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1533 CCCTTGAAGAAACTCTTCAGTAGCTCAGGCTTAAAGAAGCTGTCTGGAAGAAGCAGAAG 1592

Qy      521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGlu 540
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1593 GGGAAACGAGGAGGTGGGGGAGACGAAGAGCCTGGAGAATACCAACACATTCACACCGAA 1652

Qy      541 SerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluPro 560
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1653 TCCCCAGAGAGTGCTGATGAGCAGAAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCT 1712

Qy      561 GluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGlu 580
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1713 GAGGAGACCACGTGTCTGGAGAAAGGGCCGCTGGAAGCACCCCAGGATGGGGAAGCTGAG 1772

Qy      581 GluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPhe 600
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1773 GAAGGAACTACTTCCGATGGAGAGAAGAAGAGAGAAGGGATCACTCCCTGGGCATCCTTC 1832

Qy      601 LysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGlu 620
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1833 AAAAAGATGGTGACACCCAAGAAACGGGTCCGAAGACCTTCTGAGAGTGACAAGGAGGAA 1892

Qy      621 GluLeuGluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1893 GAGCTGGAGAAGGTCAAGAGCGCCACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATG 1952

Qy      641 GlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArgVal 660
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1953 CAAGATGAAGTCAAAACTGTTGGTGAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTG 2012

Qy      661 AspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArg 680
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2013 GATACTTCAGTGTCTTGGGAAGCACTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGG 2072

Qy      681 LysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArg 700
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2073 AAGGCATCCTCTTCAGATGATGAAGGAGGGCCAAGGACACTGGGAGGGGACAGTCACAGA 2132
```

```
Qy    701 AlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGln 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2133 GCAGAGGAGGCCAGCAAAGACAAAGAAGCCGGAACAGACGCTGTTCCTGCCAGCACCCAG 2192

Qy    721 GluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGly 740
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2193 GAGCAGGACCAAGCGCAAGGAAGTTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGG 2252

Qy    741 GluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLys 760
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2253 GAAGGTGTCTCCACTTGGGAGTCATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAG 2312

Qy    761 SerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLeuSerThrGluIleGlu 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2313 TCAAAACTGGAAGAGAAAGCCGAAGACTCTAGTGTAGAGCAGTTGTCCACTGAGATCGAA 2372

Qy    781 ProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLys 800
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2373 CCGAGTAGAGAAGAATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAA 2432

Qy    801 ArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsn 820
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2433 AGGGCAGACGGGAAGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAAT 2492

Qy    821 GluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaValGluArg 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2493 GAGGACGACCCTAATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATAATGCAGTGGAGAGG 2552

Qy    841 GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVal 860
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2553 GAGAAGATGGAAGCCCAGGGGAATACGGAGCTGCCCCAGCTGCTGGGGGCTGTGTACGTG 2612

Qy    861 SerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyThr 880
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2613 TCCGAGGAGCTCAGTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACC 2672

Qy    881 ArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGlu 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2673 AGGGCAGTCACCAGTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAA 2732

Qy    901 ProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAsp 920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2733 CCTCTTGAACACACAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGAC 2792

Qy    921 IleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHis 940
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2793 ATCATTGCAGAAGAAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCAT 2852

Qy    941 AspAspMetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThr 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2853 GACGACATGGTCACCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACC 2912

Qy    961 SerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAsp 980
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2913 TCAGAGGCTCTCCGTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGAC 2972

Qy    981 MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrPro 1000
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2973 ATGGTGTCCGCAGTTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCA 3032

Qy   1001 ValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAla 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3033 GTTCAGGAGGTAGAGAGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCC 3092

Qy   1021 IleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThr 1040
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3093 ATCCTCCAAGCCGTTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACT 3152

Qy   1041 ValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGlu 1060
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3153 GTGCAGAGAACGGGGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAA 3212

Qy   1061 ValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGly 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3213 GTGCTGGCTTCGGAGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGA 3272

Qy   1081 AlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVal 1100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3273 GCTGAGCATCTTGCACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTT 3332

Qy   1101 ProGluValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGln 1120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3333 CCTGAAGTCACGGCAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAG 3392

Qy   1121 LeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsn 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3393 CTGATGGAACAGGCCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAAT 3452

Qy   1141 GlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIle 1160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3453 GGAAGCACTCCCTTAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATT 3512

Qy   1161 AspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu 1180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3513 GACAGCCAGGACAGTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAG 3572

Qy   1181 AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGlu 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3573 GCGGCTACTGCTCAGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAA 3632

Qy   1201 GluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAla 1220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3633 GAACATGGGGAAGAACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTACTGCT 3692

Qy   1221 AlaAlaValProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu 1240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3693 GCAGCCGTGCCCGTTCTGGCAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTG 3752

Qy   1241 AspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGln 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3753 GATGGAGAAAAAGTCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAA 3812

Qy   1261 LysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLys 1280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3813 AAGGCTGCTGATGTGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAG 3872

Qy   1281 GluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGlu 1300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3873 GAGAGTACTGAAGTGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAA 3932

Qy   1301 LysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAla 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3933 AAGGAGAAAAGGGAGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCC 3992

Qy   1321 AlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSer 1340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3993 GCTCCTGAGCATGAAGGAACCTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCA 4052

Qy   1341 GluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLys 1360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4053 GAGAGGGGGAAGGCACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAA 4112

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4113 GCAGGTTGCATAGAGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAA 4172

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4173 GCTGTGGAAAAGGTCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGT 4232

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeu 1420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   4233 GTAGGTGCACACTTATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTT 4292

Qy   1421 GlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIle 1440
```

```
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4293 CAGCATGCAGAGGACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATC 4352

Qy    1441 IleValThrProAlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAla 1460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4353 ATAGTAACTCCTGCTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCA 4412

Qy    1461 SerGlnArgGluArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGly 1480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4413 TCCCAGAGAGAGCGATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGC 4472

Qy    1481 LysGluSerThrAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGlu 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4473 AAGGAGAGTACAGCAATCGAAAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAG 4532

Qy    1501 SerLysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg 1520
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4533 AGTAAGAGCAACAAGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGT 4592

Qy    1521 ThrGluThrAlaProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArg 1540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4593 ACAGAAACAGCCCCCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCATGCAGG 4652

Qy    1541 LeuAspSerArgGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHis 1560
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4653 CTTGACAGCAGGGAGCCCAACAGATGCTGGACAAAAATGAAAGATGCCAAGATGAAACAC 4712

Qy    1561 ProValProGlnProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGln 1580
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    4713 CCAGTGCCGCAGCCCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGGCTCAG 4772

Qy    1581 ProArgLysCysLeuProArgLeuGlnLeuLysAlaProValSerLys 1596
             ||||||||||||||||||||||||||||||||||||||||||||||||
Db    4773 CCTCGGAAATGCTTGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 4820
```

```
RESULT 2
US-09-902-432-1
; Sequence 1, Application US/09902432
; Patent No. US20020160002A1
; GENERAL INFORMATION:
;  APPLICANT: Irwin H. Gelman
;  APPLICANT:  Susan G. Jaken
;  TITLE OF INVENTION: TUMOR SUPPRESSOR GENE
;  FILE REFERENCE: A30558-A-FWC-A  070156.0597
;  CURRENT APPLICATION NUMBER: US/09/902,432
;  CURRENT FILING DATE:  2002-04-08
;  PRIOR APPLICATION NUMBER: 08/978,277
;  PRIOR FILING DATE: 1997-11-25
;  PRIOR APPLICATION NUMBER: 08/665,401
;  PRIOR FILING DATE: 1996-06-18
;  PRIOR APPLICATION NUMBER: 08/635,121
;  PRIOR FILING DATE: 1996-04-19
;  NUMBER OF SEQ ID NOS: 35
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 5134
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
US-09-902-432-1

Alignment Scores:
Pred. No.:                0           Length:         5134
Score:                    6005.00     Matches:        1232
Percent Similarity:       97.33%      Conservative:   7
Best Local Similarity:    96.78%      Mismatches:     27
Query Match:              74.38%      Indels:         14
DB:                       9           Gaps:           4

US-09-902-432-4 (1-1596) x US-09-902-432-1 (1-5134)

Qy     329 GluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaArg 348
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2 GAAAAGACAGAGCCAGCCTCGGAGGAGCAGGAGCCGGCAGAAGACACAGACCAGGCCAGG 61

Qy     349 LeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGlu 368
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      62 TTGTCAGCAGACTACGAGAAGGTGGAGCTGCCTTTGGAAGACCAGGTTGGTGACCTGGAG 121

Qy     369 AlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLysMetGlu 388
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     122 GCATCGTCAGAGGAGAAGTGTGCTCCTTTGGCAACGGAAGTGTTTGATGAGAAGATGGAA 181

Qy     389 AlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGlu 408
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     182 GCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAGACAGAGGAGGAG 241

Qy     409 GlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeu 428
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     242 CAGGGAGGAGGAGGAGAGGCTGAAGGGGGCGTGGTGGTAGAAGGAACAGGAGAATCCTTG 301

Qy     429 ProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGlu 448
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     302 CCCCCTGAGAAACTGGCTGAGCCCCAGGAGGTCCCCCAGGAAGCTGAGCCTGCTGAGGAG 361

Qy     449 LeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeu 468
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     362 CTGATGAAGAGCAGAGAGATGTGTGTCTCTGGAGGAGACCACACTCAACTGACAGACCTA 421

Qy     469 SerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMet 488
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     422 AGTCCTGAAGAGAAGACGCTGCCCAAACACCCAGAAGGCATTGTCAGTGAGGTGGAGATG 481

Qy     489 LeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSer 508
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     482 CTGTCCTCTCAGGAAAGAATCAAGGTACAGGGAAGTCCCTTGAAGAAACTCTTCAGTAGC 541

Qy     509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     542 TCAGGCTTAAAGAAGCTGTCTGGGAAGAAGCAGAAGGGGAAACGAGGAGGTGGGGGAGAC 601

Qy     529 GluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln 548
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     602 GAAGAGCCTGGAGAATACCAACACATTCACACCGAATCCCCAGAGAGTGCTGATGAGCAG 661

Qy     549 LysGlyGluSerSerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLys 568
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     662 AAGGGAGAGAGCTCTGCGTCGTCCCCCGAGGAGCCTGAGGAGACCACGTGTCTGGAGAAA 721

Qy     569 GlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGlu 588
             |||||||||||||||||    ||||||||||||||||||||||||||||     |||||
Db     722 GGGCCGCTGGAAGCACCCAG-GATGGGGAAGCTGAGGAAGGAACTACTTCGT--GGAGAG 778

Qy     589 LysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLys 608
             |||||||||:::   ||||||||||||||||||||||||||||||||||||||||||||
Db     779 AAGAAGAGGAAGG--ATCACTCCCTGGGCATCCTTCAAAAAGATGGTGACACCCAAGAAA 836

Qy     609 ArgValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAla 628
             |||   ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     837 CGGTC-CGAAGACCTTCTGAGAGTGACAAGGAGGAAGAGCTGGAGAAGGTCAAGAGCGCC 895

Qy     629 ThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGly 648
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     896 ACCTTGTCCTCCACTGATAGCACAGTGTCAGAAATGCAAGATGAAGTCAAAACTGTTGGT 955

Qy     649 GluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAla 668
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     956 GAGGAACAAAAGCCAGAGGAACCAAAGCGTAGGGTGGATACTTCAGTGTCTTGGGAAGCA 1015

Qy     669 LeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGlu 688
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1016 CTGATTTGTGTCGGATCATCCAAGAAGAGAGCAAGGAAGGCATCCTCTTCAGATATAAGA 1075

Qy     689 GlyGlyProArgThrLeuGly-GlyAspSerHisArgAlaGluGluAlaSerLysAspLy 708
                |||||||||||||||||| |||||||||||||||||||||||||||||||||||||
Db    1076 ---GGGCCAAGGACACTGGGAGGGGGACAGTCACAGAGCAGAGGAGGCCAGCAAAGACAA 1132
```

```
Qy    708 sGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySe 728
          |||||||   ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133 AGAAGCCGA-ACAGACGCTGTTCCTGCCAGCACCCAGGAGCAGGACCAAGCGCAAGGAAG 1191

Qy    728 rSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSe 748
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1192 TTCCTCACCCGAGCCAGCGGGAAGCCCTTCCGAAGGGGAAGGTGTCTCCACTTGGGAGTC 1251

Qy    748 rPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAla-- 767
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1252 ATTTAAAAGATTAGTCACTCCAAGAAAAAAATCCAAGTCAAAACTGGAAGAGAAAGA-AG 1310

Qy    768 ----GluAspSerSerVal---GluGlnLeuSerThrGluIleGluProSerArgGluGl 785
              :::|||||||||      :::||||||||||||||||||||   |||||||
Db   1311 CCGGAAGGACTCTAGTTGTAGGAGCAGGTTGTCCACTGAGATCGAACCGTGTAGAGAAGA 1370

Qy    785 uSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLy 805
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1371 ATCTTGGGTTTCCATTAAGAAATTCATCCCCGGACGGCGGAAGAAAAGGGCAGATGGGAA 1430

Qy    805 s-GlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProA 825
          | |||||||||||||||||||||||||||||||||||||||||||||||||:::|||:
Db   1431 GGCAAGAACAAGCCACTGTGGAAGACTCAGGGCCAGTGGAGATAAATGAGGACGAGCCTG 1490

Qy    825 snValProAlaValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluA 845
          ::||||||||||||||||||||||||||||||:::||||||||||||||||||||||||
Db   1491 ATGTCCCAGCCGTCGTGCCTCTGTCTGAGTATGATGCAGTGGAGAGGGAGAAGATGGAAG 1550

Qy    845 laGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuS 865
          ||||||||||. ||||||||         |||  |||   ||||||||||||||||
Db   1551 CCCAGGGGAATGCGGAGCTGCCCAGCTGCTGGGGCTGTGTA---GTGTCCGAGGAGCTCA 1607

Qy    865 erLysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrS 885
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1608 GTAAGACTCTGGTCCACACTGTGAGTGTCGCAGTCATTGATGGGACCAGGGCAGTCACCA 1667

Qy    885 erValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisT 905
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1668 GTGTCGAAGAGCGGTCTCCTTCGTGGATATCCGCTTCCGTAACAGAACCTCTTGAACACA 1727

Qy    905 hrAlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluG 925
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1728 CAGCGGGAGAAGCCATGCCACCTGTTGAAGAGGTCACTGAAAAAGACATCATTGCAGAAG 1787

Qy    925 luThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetValT 945
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1788 AAACTCCTGTGCTCACCCAGACGTTACCAGAGGGTAAAGATGCCCATGACGACATGGTCA 1847

Qy    945 hrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuA 965
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1848 CCAGTGAAGTGGATTTCACCTCAGAAGCTGTGACAGCCACAGAGACCTCAGAGGCTCTCC 1907

Qy    965 rgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaV 985
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1908 GTACTGAAGAAGTTACCGAAGCATCGGGGGCCGAAGAGACCACAGACATGGTGTCCGCAG 1967

Qy    985 alSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValG 1005
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1968 TTTCCCAGCTGACTGACTCCCCAGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAG 2027

Qy   1005 luSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaV 1025
          ||   |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2028 AGGGTGGTGTGCTAGATACAGAAGAAGAGGAGCGCCAGACGCAGGCCATCCTCCAAGCCG 2087

Qy   1025 alAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrG 1045
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2088 TTGCAGACAAGGTGAAAGAGGAGTCCCAGGTGCCTGCAACCCAGACTGTGCAGAGAACGG 2147

Qy   1045 lySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerG 1065
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2148 GGTCAAAAGCACTGGAGAAGGTTGAGGAGGTAGAGGAGGACTCCGAAGTGCTGGCTTCGG 2207

Qy   1065 luLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuA 1085
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2208 AGAAAGAGAAGGACGTTATGCCGAAAGGACCCGTGCAGGAAGCTGGAGCTGAGCATCTTG 2267

Qy   1085 laGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrA 1105
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2268 CACAGGGCTCTGAGACTGGACAGGCTACTCCAGAGAGCCTTGAAGTTCCTGAAGTCACAG 2327

Qy   1105 laAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnA 1125
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2328 CAGATGTAGACCATGTCGCCACGTGCCAGGTTATCAAGCTCCAGCAGCTGATGGAACAGG 2387

Qy   1125 laValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProL 1145
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2388 CCGTGGCCCCTGAGTCATCCGAAACCTTGACAGACAGTGAGACAAATGGAAGCACTCCCT 2447

Qy   1145 euAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspS 1165
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2448 TAGCAGATTCAGACACTGCAGATGGGACACAGCAAGATGAAACCATTGACAGCCAGGACA 2507

Qy   1165 erLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaG 1185
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2508 GTAAAGCCACTGCAGCTGTCAGGCAGTCACAGGTCACAGAAGAAGAGGCGGCTACTGCTC 2567

Qy   1185 lnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluG 1205
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2568 AGAAAGAGGAGCCTTCGACACTACCTAATAATGTTCCAGCCCAGGAAGAACATGGGGAAG 2627

Qy   1205 luProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProV 1225
          ||||||||||||||||||||||||||||||||||||   ||||||||||||||||||
Db   2628 AACCAGGAAGAGATGTTCTTGAACCTACACAGCAAGAGCTTGCTGCTGCAGCCGTGCCCG 2687

Qy   1225 alLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysV 1245
          ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2688 TCTGGCAAAAGACTGAGGTGGGTCAAGAGGGTGAGGTTGACTGGTTGGATGGAGAAAAAG 2747

Qy   1245 alLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspV 1265
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2748 TCAAAGAAGAACAGGAGGTGTTTGTACACTCTGGACCCAACAGTCAAAAGGCTGCTGATG 2807

Qy   1265 alThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluV 1285
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2808 TGACATATGACAGTGAAGTGATGGGAGTGGCCGGGTGTCAGGAAAAGGAGAGTACTGAAG 2867

Qy   1285 alGlnSerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgG 1305
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2868 TGCAGAGTCTTAGCCTGGAGGAGGGAGAGATGGAAACTGACGTTGAAAAGGAGAAAAGGG 2927

Qy   1305 luThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisG 1325
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2928 AGACAAAGCCAGAGCAAGTGAGTGAAGAAGGTGAGCAGGAAACAGCCGCTCCTGAGCATG 2987

Qy   1325 luGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysA 1345
          ||      ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2988 AAAGGAACTACGGGAAGCCAGTCCTGACACTTGACATGCCCAGCTCAGAGAGGGGGAAGG 3047

Qy   1345 laLeuGlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleG 1365
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3048 CACTGGGAAGCCTTGGAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGGTTGCATAG 3107

Qy   1365 luValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysV 1385
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3108 AGGTTCAAGTTCAAAGCCTGGACACAACAGTCACTCAAACAGCAGAAGCTGTGGAAAAGG 3167

Qy   1385 alIleGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisL 1405
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3168 TCATAGAAACGGTTGTGATTTCAGAGACAGGTGAAAGTCCAGAGTGTGTAGGTGCACACT 3227

Qy   1405 euLeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluA 1425
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   3228 TATTACCAGCTGAGAAGTCCTCTGCAACGGGTGGCCACTGGACTCTTCAGCATGCAGAGG 3287

Qy   1425 spThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrProA 1445
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db   3288 ACACGGTACCCCTGGGGCCTGAGTCTCAGGCAGAATCCATCCCAATCATAGTAACTCCTG 3347

Qy   1445 laProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluA 1465
Db   3348 CTCCTGAAAGCACCCTACATCCTGACCTACAAGGAGAAATAAGCGCATCCCAGAGAGAGC 3407

Qy   1465 rgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrA 1485
Db   3408 GATCAGAGGAAGAGGACAAGCCAGATGCTGGTCCTGATGCTGACGGCAAGGAGAGTACAG 3467

Qy   1485 laIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnL 1505
Db   3468 CAATCGACAAAGTCCTCAAGGCTGAACCTGAGATCCTGGAACTTGAGAGTAAGAGCAACA 3527

Qy   1505 ysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaP 1525
Db   3528 AGATTGTGCTGAACGTCATTCAGACAGCCGTTGACCAGTTCGCACGTACAGAAACAGCCC 3587

Qy   1525 roGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgG 1545
Db   3588 CCGAAACTCATGCTTATGATTCACAGACCCAGGTTCCTGCAATGCGCTTGGACAGCAGGG 3647

Qy   1545 luProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnP 1565
Db   3648 AGCCCAACAGATGCTGGACAAAAATGAAAGTTGCCAAGATGAAACACCCAGTGCCGCAGC 3707

Qy   1565 roArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysL 1585
Db   3708 CCAGAGAGGACTTGCAAGTCCTGACCGTTCTGGAGGCATGGC--TCAGCTCGGAAATGCT 3765

Qy   1585 euProArgLeuGlnLeuLysAlaProValSerLys 1596
Db   3766 TGCCGCGCTTGCAGTTGAAAGCGCCGGTGTCAAAG 3800

RESULT 3
US-09-738-877-2
; Sequence 2, Application US/09738877
; Patent No. US20020015970A1
; GENERAL INFORMATION:
;  APPLICANT: Murray, Richard
;  APPLICANT:  Watson, Susan
;  APPLICANT:  Weiss, Stephen J.
;  APPLICANT:  Glynne, Richard
;  TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET
;  TITLE OF INVENTION:  SCREENING FOR ANGIOGENESIS MODULATORS
;  FILE REFERENCE: A-69806/DJB/JJD
;  CURRENT APPLICATION NUMBER: US/09/738,877
;  CURRENT FILING DATE:  2000-12-15
;  PRIOR APPLICATION NUMBER: US 09/637,977
;  PRIOR FILING DATE: 2000-08-11
;  PRIOR APPLICATION NUMBER: US 60/148,425
;  PRIOR FILING DATE: 1999-08-11
;  PRIOR APPLICATION NUMBER: PCT/US 00/22061
;  PRIOR FILING DATE: 2000-08-11
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 5346
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-738-877-2

Alignment Scores:
Pred. No.:              1.22e-292     Length:         5346
Score:                  4300.50       Matches:        967
Percent Similarity:     68.81%        Conservative:   189
Best Local Similarity:  57.56%        Mismatches:     408
Query Match:            53.27%        Indels:         116
DB:                     10            Gaps:           36

US-09-902-432-4 (1-1596) x US-09-738-877-2 (1-5346)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db      1 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 60

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db     61 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 120

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    121 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 180

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    181 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 240

Qy     77 GlnGluGlyGln----------------------------------------GluGlu 82
Db    241 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 300

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    301 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 360

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    361 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 420

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    421 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 477

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    478 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 537

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    538 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 597

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    598 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 648

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    649 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 702

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    703 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 762

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    763 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 819

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db    820 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 879

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db    880 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 939

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db    940 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 999

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1000 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1059

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
```

```
Db   1060 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1119

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1120 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1179

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1180 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1239

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1240 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1290

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1291 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1335

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1336 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1395

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1396 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1455

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1456 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1515

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1516 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1575

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1576 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1632

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1633 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1692

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1693 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1752

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1753 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 1812

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   1813 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 1872

Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   1873 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 1932

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   1933 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 1992

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   1993 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2052

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2053 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2112

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2113 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2172

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2173 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2232

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2233 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2292

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2293 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2352

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2353 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2412

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2413 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2472

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2473 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2532

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2533 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2592

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2593 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2652

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2653 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2712

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2713 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2772

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2773 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 2832

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   2833 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 2892

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   2893 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 2952

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   2953 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3012

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3013 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3072

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3073 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3132

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3133 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3192

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3193 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3234
```

```
Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
             :::|||          :::||||||  ||||||:::    |||          |||
Db   3235 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3294

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
          ||||||      |||        ||||||            ||||||  ||||||||||||
Db   3295 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3354

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
           |||      |||:::||||||       ::::::             |||   ||||||
Db   3355 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3414

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
          |||                  :::|||  |||::::::   ||||||||||||:::
Db   3415 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3474

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
          |||:::|||  ||||||  ||||||||||||||:::||||||||||||:::||||||
Db   3475 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3534

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
             |||         ||||||:::||||||   ::::::       :::
Db   3535 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3594

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGlu 1187
          |||:::       ||||||      ||||||  ||||||        ||||||||||||
Db   3595 GTCGCATCTGGTACCCAGTCAGGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3654

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
             |||       |||::::::       |||||||||      |||:::
Db   3655 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3711

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaValProValLeu 1226
                |||  ||||||  |||     :::|||::::::          |||    :::|||
Db   3712 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3771

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
          :::|||||||||      |||      |||  |||                ||||||   |||
Db   3772 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 3825

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
                          |||    |||       :::  :::|||
Db   3826 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 3885

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
          :::|||       :::|||               |||         |||  |||::::::
Db   3886 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 3945

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
             ::::::   |||:::|||||                  |||        |||   |||
Db   3946 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4005

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
          |||         ||||||:::||||||  |||   :::|||       |||:::||||||
Db   4006 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4065

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
                |||   |||||||||                  |||          :::   :::
Db   4066 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4125

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
          |||::::::::|||       :::      |||   :::   ||||||   ||||||
Db   4126 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4185

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
          |||          |||::::::|||   |||   ::::::|||||||||||||    :::
Db   4186 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4245

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
             :::   |||   ||||||   |||   |||:::   :::      ||||||
Db   4246 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4305

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
          |||   ||||||||||||:::   |||      ||||||||||||:::      ||||||
Db   4306 ATTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4365

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
          ||||||                  :::                  ||||||  ||||||  |||
Db   4366 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4425

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
          |||:::   ||||||:::|||   |||:::||||||:::            :::      |||
Db   4426 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4485

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
             ||||||:::||||||   :::   |||   :::   :::|||:::|||      |||
Db   4486 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4545

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
          :::      |||      :::         |||   |||   |||||||||      :::
Db   4546 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4596

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
                |||||||||      ||||||||||||||:::||||||:::|||:::|||
Db   4597 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4650

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
          |||:::||||||||||||||||||||||||   ||||||   |||||||||   |||
Db   4651 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4710

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
             :::   |||||||||               :::   ||||||::::::
Db   4711 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4770

Qy   1548 ArgCysTrpThrLys--------------MetLysAspAla-LysMetLysHisProVa 1562
          :::      |||                      |||   :::|||||||||      ::
Db   4771 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 4830

Qy   1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
          :   ||||||:::                        |||      :::|||
Db   4831 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 4890

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
            :::         :::|||:::|||          |||:::   |||   |||:::
Db   4891 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 4948
```

RESULT 4
US-09-738-877-1
; Sequence 1, Application US/09738877
; Patent No. US20020015970A1
; GENERAL INFORMATION:
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan
; APPLICANT: Weiss, Stephen J.
; APPLICANT: Glynne, Richard
; TITLE OF INVENTION: NOVEL METHODS OF DIAGNOSIS OF ANGIOGENESIS, COMPOSITIONS, AND MET
; TITLE OF INVENTION: SCREENING FOR ANGIOGENESIS MODULATORS
; FILE REFERENCE: A-69806/DJB/JJD
; CURRENT APPLICATION NUMBER: US/09/738,877
; CURRENT FILING DATE: 2000-12-15
; PRIOR APPLICATION NUMBER: US 09/637,977
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: US 60/148,425
; PRIOR FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: PCT/US 00/22061
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 6608
; TYPE: DNA
; ORGANISM: Homo sapiens
; PUBLICATION INFORMATION:
; AUTHORS: Nauert et al.
; TITLE: Gravin, an autoantigen recognized by serum from myastheinia gravis
; TITLE: patients, is a kinase scaffold protein

```
;   JOURNAL: Curr. Biol.
;   VOLUME: 7
;   ISSUE: 1
;   PAGES: 52-62
;   DATE: 1997-01-01
;   DATABASE ACCESSION NUMBER: U81607
;   DATABASE ENTRY DATE: 1997-06-26
US-09-738-877-1

Alignment Scores:
Pred. No.:                1.56e-292     Length:         6608
Score:                    4300.50       Matches:        967
Percent Similarity:       68.81%        Conservative:   189
Best Local Similarity:    57.56%        Mismatches:     408
Query Match:              53.27%        Indels:         116
DB:                       10            Gaps:           36

US-09-902-432-4 (1-1596) x US-09-738-877-1 (1-6608)

Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln------------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db    840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy    223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db    894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy    243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db    954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy    263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db   1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy    283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db   1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy    303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db   1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy    323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
Db   1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy    338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db   1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy    354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db   1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy    374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db   1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy    393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db   1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy    413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db   1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy    433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db   1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy    453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db   1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy    473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db   1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy    493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db   1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy    513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db   1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy    533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db   1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy    553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db   1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy    573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db   1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy    593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db   2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063
```

```
Qy    613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db   2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy    633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db   2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy    653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db   2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy    673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db   2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy    693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db   2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy    713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db   2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGGAAGTTCCTCCCCGGAG 2423

Qy    733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db   2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy    753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db   2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy    770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db   2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy    789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db   2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy    809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db   2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy    829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db   2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy    846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db   2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy    866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db   2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy    886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db   2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy    906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db   2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023

Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3845

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3846 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3902

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3903 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3962

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3963 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4016

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4017 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4076

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4077 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4136

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4137 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4196

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
```

```
Db   4197 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4256

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4257 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4316

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4317 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4376

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4377 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4436

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4437 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4496

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4497 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4556

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4557 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4616

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4617 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4676

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4677 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4736

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4737 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4787

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
Db   4788 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4841

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
Db   4842 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4901

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
Db   4902 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4961

Qy   1548 ArgCysTrpThrLys--------------MetLysAspAla-LysMetLysHisProVa 1562
Db   4962 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5021

Qy   1562 lProGlnProArgGlu--------------------AspLeuGlnValLeuThrValLe 1575
Db   5022 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5081

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   5082 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5139
```

RESULT 5
US-09-880-107-3439
; Sequence 3439, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3439
; LENGTH: 6608
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 U81607
US-09-880-107-3439

Alignment Scores:

Pred. No.:	1.56e-292	Length:	6608
Score:	4300.50	Matches:	967
Percent Similarity:	68.81%	Conservative:	189
Best Local Similarity:	57.56%	Mismatches:	408
Query Match:	53.27%	Indels:	116
DB:	10	Gaps:	36

US-09-902-432-4 (1-1596) x US-09-880-107-3439 (1-6608)

```
Qy      1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
Db    192 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 251

Qy     20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
Db    252 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 311

Qy     39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
Db    312 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 371

Qy     57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
Db    372 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 431

Qy     77 GlnGluGlyGln------------------------------------------GluGlu 82
Db    432 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 491

Qy     83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
Db    492 GAAGTCATTGTCACGGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCGAAAGAGACTCC 551

Qy    103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
Db    552 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 611

Qy    123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
Db    612 GAGAACCGAAAT---ATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 668

Qy    143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
Db    669 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 728

Qy    163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
Db    729 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 788

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db    789 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGAC 839

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
```

```
Db     840 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 893

Qy     223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
Db     894 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 953

Qy     243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
Db     954 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 1010

Qy     263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
Db    1011 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 1070

Qy     283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
Db    1071 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 1130

Qy     303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
Db    1131 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1190

Qy     323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu-------------- 337
Db    1191 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1250

Qy     338 -----------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
Db    1251 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1310

Qy     354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
Db    1311 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1370

Qy     374 LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGln---Glu 392
Db    1371 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAG 1430

Qy     393 ValValAlaGluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGly 412
Db    1431 GTTGTGGCCGAAGTCCACGTCAGCACCGTGGAGGAGAGAACCGAAGAGCAG--------- 1481

Qy     413 GlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLys 432
Db    1482 ---------------AAAACGGAGGTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAA 1526

Qy     433 LeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSer 452
Db    1527 TTGGTTGGAATGGATGCAGAACCTCAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTC 1586

Qy     453 ArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGlu 472
Db    1587 AAAGAAACGTGTGTTTCCGGAGAGGACCCTACACAGGGAGCTGACCTCAGTCCTGATGAG 1646

Qy     473 LysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGln 492
Db    1647 AAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAG 1706

Qy     493 GluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLys 512
Db    1707 GAGAGAATGAAGGTGCAGGGAAGTCCACTAAAGAAGCTTTTTACCAGCACTGGCTTAAAA 1766

Qy     513 LysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAspGluGluProGly 532
Db    1767 AAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA---GGAGGAGGAGACGAGGAATCAGGG 1823

Qy     533 GluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSer 552
Db    1824 GAGCACACTCAGGTTCCAGCCGATTCTCCGGACAGCCAGGAGGAGCAAAAGGGCGAGAGC 1883

Qy     553 SerAlaSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
Db    1884 TCTGCCTCATCCCCTGAGGAGCCCGAGGAGATCACGTGTCTGGAAAAGGGCTTAGCCGAG 1943

Qy     573 AlaProGlnAspGlyGluAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGlu 592
Db    1944 GTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCTACTTCCGATGGAGAGAAAAAAAGAGAA 2003

Qy     593 GlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValArgArg 612
Db    2004 GGTGTCACTCCCTGGGCATCATTCAAAAAGATGGTGACGCCCAAGAAGCGTGTTAGACGG 2063

Qy     613 ProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSer 632
Db    2064 CCTTCGGAAAGTGATAAAGAAGATGAGCTGGACAAGGTCAAGAGCGCTACCTTGTCTTCC 2123

Qy     633 ThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGluGluGlnLys 652
Db    2124 ACCGAGAGCACAGCCTCTGAAATGCAAGAAGAAATGAAAGGGAGCGTGGAAGAGCCAAAG 2183

Qy     653 ProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeuIleCysVal 672
Db    2184 CCGGAAGAACCAAAGCGCAAGGTGGATACCTCAGTATCTTGGGAAGCTTTAATTTGTGTG 2243

Qy     673 GlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArg 692
Db    2244 GGATCATCCAAGAAAAGAGCAAGGAGAAGGTCCTCTTCTGATGAGGAAGGGGGACCAAAA 2303

Qy     693 ThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThr 712
Db    2304 GCAATGGGAGGAGACCACCAGAAAGCTGATGAGGCCGGAAAAGACAAAGAGACGGGGACA 2363

Qy     713 AspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGlu 732
Db    2364 GACGGGATCCTTGCTGGTTCCCAAGAACATGATCCAGGGCAGGAAGTTCCTCCCCGGAG 2423

Qy     733 ProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeu 752
Db    2424 CAAGCTGGAAGCCCTACCGAAGGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTA 2483

Qy     753 ValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAsp--------- 769
Db    2484 GTCACGCCAAGAAAAAAATCAAAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCT 2543

Qy     770 ---SerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpVal 788
Db    2544 GGGTCTGGTGTAGAACATTCCACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTC 2603

Qy     789 SerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGln 808
Db    2604 TCAATCAAGAAGTTTATTCCTGGACGAAGGAAGAAAAAGGCCAGATGGGAAACAAGAACAA 2663

Qy     809 AlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAla 828
Db    2664 GCCCCTGTTGAAGACGCAGGGCCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCC 2723

Qy     829 ValValProLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---------Ala 845
Db    2724 GTGGTCCCTCTGTCTGAGTATGATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCC 2783

Qy     846 GlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSer 865
Db    2784 CAAAAAGGCGCAGAGCAGCCCGAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGC 2843

Qy     866 LysThrLeuValHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSer 885
Db    2844 GAGAGTCAGGTTCATATGATGGCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATT 2903

Qy     886 ValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGluHisThr 905
Db    2904 ATTGAAGAAAGGTCTCCTTCTTGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTA 2963

Qy     906 AlaGlyGluAlaMetProProValGluGluValThrGluLysAspIleIleAla---Glu 924
Db    2964 GAAGCTGAAGCCGCACTGTTAACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAA 3023
```

```
Qy    925 GluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAspMetVal 944
Db   3024 GAACCCCCCACGGTTACTGAACCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTC 3083

Qy    945 ThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeu 964
Db   3084 GTTAGTGAGGCGGAATTGACCCCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGCCATTG 3143

Qy    965 ArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAla 984
Db   3144 GGTTCCGAAGAAGGAACCGAAGCATCTGCTGCTGAAGAGACCACAGAAATGGTGTCAGCA 3203

Qy    985 ValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVal 1004
Db   3204 GTCTCCCAGTTAACCGACTCCCCAGACACCACAGAGGAGGCCACTCCGGTGCAGGAGGTG 3263

Qy   1005 GluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
Db   3264 GAAGGTGGCGTACCTGACATAGAAGAGCAAGAGAGGCGGACTCAAGAGGTCCTCCAGGCA 3323

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThr------------------ 1038
Db   3324 GTGGCAGAAAAAGTGAAAGAGGAATCCCAGCTGCCTGGCACCGGTGGGCCAGAAGATGTG 3383

Qy   1039 ---GlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluValGluGlu 1057
Db   3384 CTTCAGCCTGTGCAGAGAGCA------------------GAGGCAGAAAGACCAGAAGAG 3425

Qy   1058 AspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProValGln 1077
Db   3426 CAGGCTGAAGCGTCGGGTCTGAAGAAAGAGACGGATGTAGTGTTGAAAGTAGATGCTCAG 3485

Qy   1078 GluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSer 1097
Db   3486 GAGGCAAAAACTGAGCCTTTTACACAAGGGAAGGTGGTGGGGCAGACCACCCCAGAAAGC 3545

Qy   1098 LeuGlu---ValProGluValThrAlaAspValAspHis---------ValAlaThrCys 1113
Db   3546 TTTGAAAAAAGCTCCTCAAGTCACAGAGAGCATAGAGTCCAGTGAGCTTGTAACCACTTGT 3605

Qy   1114 GlnVal---------------IleLysLeuGlnGlnLeu---MetGluGlnAlaValAla 1127
Db   3606 CAAGCCGAAACCTTAGCTGGGGTAAAATCACAGGAGATGGTGATGGAACAGGCTATCCCC 3665

Qy   1128 ProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAsp 1147
Db   3666 CCTGACTCGGTGGAAACCCCTACAGACAGTGAGACTGATGGAAGCACCCCCGTAGCCGAC 3725

Qy   1148 SerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLysAla 1167
Db   3726 TTTGACGCACCAGGCACAACCCAGAAAGACGAGATTGTGGAAATCCATGAGGAGAATGAG 3785

Qy   1168 ThrAlaAlaValArgGlnSerGlnValThrGluGluAlaAlaThrAlaGlnLysGlu 1187
Db   3786 GTCGCATCTGGTACCCAGTCAGGGGGCACAGAAGCAGAGGCAGTTCCTGCACAGAAAGAG 3845

Qy   1188 GluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGly 1207
Db   3846 AGGCCTCCAGCA---CCTTCCAGTTTTGTGTTCCAGGAAGAAACTAAAGAACAATCAAAG 3902

Qy   1208 ---ArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeu 1226
Db   3903 ATGGAAGACACTCTAGAGCATACAGATAAAGAGGTGTCAGTGGAAACTGTATCCATTCTG 3962

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLys 1246
Db   3963 TCAAAGACTGAGGGGACTCAA------GAGGCTGACCAGTATGCTGATGAGAAAACCAAA 4016

Qy   1247 ---------------GluGluGlnGluValPheValHisSerGly------ProAsnSer 1259
Db   4017 GACGTACCATTTTTCGAAGGACTTGAGGGGTCTATAGACACAGGCATAACAGTCAGTCGG 4076

Qy   1260 GlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysGlnGlu 1279
Db   4077 GAAAAGGTCACTGAAGTTGCCCTTAAAGGTGAAGGGACAGAAGAAGCTGAATGTAAAAAG 4136

Qy   1280 LysGluSerThrGluValGlnSer---------------LeuSerLeuGluGluGlyGlu 1294
Db   4137 GATGATGCTCTTGAACTGCAGAGTCACGCTAAGTCTCCTCCATCCCCCGTGGAGAGAGAG 4196

Qy   1295 MetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu 1314
Db   4197 ATGGTAGTTCAAGTCGAAAGGGAGAAAACAGAAGCAGAGCCAACCCATGTGAATGAAGAG 4256

Qy   1315 Gly---GluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeu 1333
Db   4257 AAGCTTGAGCACGAAACAGCTGTTACCGTATCTGAAGAGGTCAGTAAGCAGCTCCTCCAG 4316

Qy   1334 ThrLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySer--- 1352
Db   4317 ACAGTGAATGTGCCCATCATAGATGGGGCAAAGGAAGTCAGCAGTTTGGAAGGAAGCCCT 4376

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
Db   4377 CCTCCCTGCCTAGGTCAAGAGGAGGCAGTATGCACCAAAATTCAAGTTCAGAGCTCTGAG 4436

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIle------GluThrValVal 1390
Db   4437 GCATCATTCACTCTAACAGCGGCTGCAGAGGAGGAAAAGGTCTTAGGAGAAACTGCCAAC 4496

Qy   1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
Db   4497 ATTTTAGAAACAGGTGAAACGTTGGAGCCTGCAGGTGCACATTTAGTTCTGGAAGAGAAA 4556

Qy   1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
Db   4557 TCCTCTGAAAAAAATGAAGACTTTGCCGCTCATCCAGGGGAAGATGCTGTGCCCACAGGG 4616

Qy   1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
Db   4617 CCCGACTGTCAGGCAAAATCGACACCAGTGATAGTATCTGCTACTACCAAGAAAGGCTTA 4676

Qy   1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
Db   4677 AGTTCCGACCTGGAAGGAGAGAAAACCACATCACTGAAGTGGAAGTCAGATGAAGTCGAT 4736

Qy   1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
Db   4737 GAGCAGGTTGCTTGCCAGGAGGTC------AAAGTGAGTGTAGCAATTGAGGATTTA--- 4787

Qy   1491 LysAlaGluProGlu------IleLeuGluLeuGluSerLysSerAsnLysIleValLeu 1508
Db   4788 ------GAGCCTGAAAATGGGATTTTGGAACTTGAGACCAAAAGCAGTAAACTTGTCCAA 4841

Qy   1509 AsnValIleGlnThrAlaValAspGlnPheAlaArgThr---GluThrAlaProGluThr 1527
Db   4842 AACATCATCCAGACAGCCGTTGACCAGTTTGTACGTACAGAAGAAACAGCCACCGAAATG 4901

Qy   1528 HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsn 1547
Db   4902 TTGACGTCTGAGTTACAGACACAAGCTCACGTGATAAAAGCTGACAGCCAGGACGCTGGA 4961

Qy   1548 ArgCysTrpThrLys---------------MetLysAspAla-LysMetLysHisProVa 1562
Db   4962 CAGGAAACGGAGAAAGAAGGAGAGGAACCTCAGGCCTCTGCACAGGATGAAACACCAATT 5021

Qy   1562 lProGlnProArgGlu---------------------AspLeuGlnValLeuThrValLe 1575
Db   5022 ACTTCAGCCAAAGAGGAGTCAGAGTCAACCGCAGTGGGACAAGCACATTCTGATATTTCC 5081

Qy   1575 uGluAlaTrpAlaGlnProArgLysCysLeuProArgLeuGlnLeuLysAlaProVal 1594
Db   5082 AAAGACATGAGTGAAGCCTCAGAAAAGACCATGACTGTTGAGGTAGAAGGTTCCACTG 5139
```

RESULT 6
US-09-742-096-2

```
; Sequence 2, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
;  APPLICANT: DRUILHE, PIERRE
;  APPLICANT:  DAUBERSIES, PIERRE
;  TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
;  FILE REFERENCE: 200773US0DIV
;  CURRENT APPLICATION NUMBER: US/09/742,096
;  CURRENT FILING DATE:  2000-12-22
;  PRIOR APPLICATION NUMBER: US 08/973,642
;  PRIOR FILING DATE: 1998-02-06
;  PRIOR APPLICATION NUMBER: PCT/FR96/00894
;  PRIOR FILING DATE: 1996-06-12
;  PRIOR APPLICATION NUMBER: FR 95/07007
;  PRIOR FILING DATE: 1995-06-13
;  NUMBER OF SEQ ID NOS: 29
;  SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 5361
;   TYPE: DNA
;   ORGANISM: P. falciparum
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(5361)
;   OTHER INFORMATION:
US-09-742-096-2

Alignment Scores:
Pred. No.:              1.68e-23      Length:        5361
Score:                  462.00        Matches:       323
Percent Similarity:     36.36%        Conservative:  293
Best Local Similarity:  19.07%        Mismatches:    646
Query Match:            5.72%         Indels:        432
DB:                     9             Gaps:          68

US-09-902-432-4 (1-1596) x US-09-742-096-2 (1-5361)

Qy      53 GlnLysAsnGlyGlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisVal 72
              |||  ||||||:::|||:::|||                 |||:::         :::
Db     283 CAAGTAAATGGTGAATTAGCTAGT---------------GAAGAAGTAAAGGAAAAAATT 327

Qy      73 GlnGluGluAsnGlnGluGlyGlnGlu---GluGluValValAspGluAspValGlyGln 91
                 :::      :::||||||              |||    ||||||::::::
Db     328 CTTGACTTATTAGAAGAAGGAAATACATTAACTGAAAGTGTAGATGATAATAAAAATTTA 387

Qy      92 ArgGluSerGluAspValArgGluLys---------AspArgValGluGlu--------- 105
              |||:::||||||:::::|||                :::||||||
Db     388 GAAGAAGCCGAAGATATAAAGGAAAATATCTTATTAAGTAATATAGAAGAACCAAAAGAA 447

Qy     106 ----------------------------MetAlaAlaAsnSerThrAlaValGluAspIle 116
                                       :::      ||||||        |||    :::
Db     448 AATATTATTGACAATTTATTAAATAATATTGGACAAAATTCAGAAAAACAAGAAAGTGTA 507

Qy     117 ThrLysAspGlyGln---------GluGluThrSerGluIleIleGluGlnIleProAla 133
           ::::::::    |||         |||     :::||||::::::          :::
Db     508 TCAGAAAATGTACAAGTCAGTGATGAACTTTTTAATGAATTATTAAATAGTGTAGATGTT 567

Qy     134 SerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGly 153
           :::                |||   :::::::::         |||||||||  ||||||
Db     568 AATGGAGAAGTAAAAGAAAATATTTTGGAG------GAAAGTCAAGTTAATGACGATATT 621

Qy     154 PheLysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGlu 173
           |||       :::   |||                       :::||||::::::::::::
Db     622 TTTAATAGTTTAGTAAAA------------------AGTGTTCAACAAGAACAACAACAC 663

Qy     174 LysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAla 193
              :::                   :::||||::::::::::::    |||      |||
Db     664 AATGTTGAAGAAAAAGTTGAAGAAAGTGTAGAAGAAAATGACGAAGAAAGTGTAGAAGAA 723

Qy     194 SerVal------GlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGlu 211
           :::|||               ::::::   :::      ||||||   ::::::|||   |||
Db     724 AATGTAGAAGAAAATGTAGAAGAAAATGACGACGGAAGTGTAGCCTCAAGTGTTGAAGAA 783

Qy     212 Ser------AlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThr 229
           |||         :::|||   :::        :::      |||    |||:::
Db     784 AGTATAGCTTCAAGTGTTGATGAAAGTATAGATTCAAGTATTGAAGAAAATGTAGCTCCA 843

Qy     230 LeuLysGlnGlu------------------------------------------------ 233
              :::|||
Db     844 ACTGTTGAAGAAATCGTAGCTCCAAGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAA 903

Qy     234 ------GlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGlu 251
                 :::   :::   |||      :::         :::::::::         |||
Db     904 AGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAA 963

Qy     252 GluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSer 271
              ::::::            |||      |||:::           ::::::   |||
Db     964 AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAA 1023

Qy     272 ProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGly 291
              :::|||            |||      :::         :::
Db    1024 ATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTGTAGCTCCA 1083

Qy     292 TrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGlu--- 310
                                |||   ::::::|||   |||:::::::::|||
Db    1084 AGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAA 1143

Qy     311 ---ThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGlu 329
                 ||||||        :::|||   ||||||:::|||            |||
Db    1144 AGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAA 1203

Qy     330 LysThrGluProAla---------SerGluGluGlnGluProAlaGluAspThrAspGln 346
              |||   :::         :::   |||:::      ||||||:::   ::::::
Db    1204 AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAA 1263

Qy     347 AlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAsp 366
              ::::::         |||::::::      |||         :::::::::
Db    1264 ATC---GTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATT------ 1314

Qy     367 LeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLys 386
           :::      |||      |||      ||||||         |||      :::|||
Db    1315 GTAGCTCCAAGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAAT 1374

Qy     387 MetGluAlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGlu 406
           :::|||    :::      |||         |||      |||      |||      |||
Db    1375 GTTGAA------GAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAA 1428

Qy     407 GluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGlu 426
           |||                                     |||   |||         |||
Db    1429 GAAAGT--------------------------------GTAGCTGAAAATGTTGAAGAA 1455

Qy     427 SerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAla 446
           |||:::                                          ||||||
Db    1456 AGTGTA------------------------------------------GCTGAAAATGTT 1473

Qy     447 GluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThr 466
           ||||||   :::
Db    1474 GAAGAAAGTGTAGCTGAA------------------------------------------ 1491

Qy     467 AspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluVal 486
              :::   ||||||               |||   :::               |||
Db    1492 ------AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTT 1545

Qy     487 GluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPhe 506
           |||   :::   :::               ::::::::::::
Db    1546 GAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTGTA 1605

Qy     507 SerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGly 526
           :::   |||   ::::::   ::::::
Db    1606 GCTCCAAGTGTTGTAGAAAGTGTGGCT--------------------------------- 1632

Qy     527 GlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAsp 546
                       |||         :::   :::               ||||||
Db    1633 ------------CCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCT 1680

Qy     547 GluGlnLysGlyGluSerSerAlaSerSerProGluGluPro---------------Glu 561
```

```
Db   1681 GAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAA 1740

Qy    562 GluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGlu 581
Db   1741 GAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCT 1800

Qy    582 GlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLys 601
Db   1801 CCAACTGTT-----------------GAAGAAATTGTAGCTCCA------AGTGTTGTA 1836

Qy    602 LysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGlu 621
Db   1837 GAAAGTGTGGCTCCAAGT------GTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGT 1890

Qy    622 Leu---GluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
Db   1891 GTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAAT 1950

Qy    641 GlnAspGluVal------LysThrValGlyGluGluGlnLysProGluGluProLysArg 658
Db   1951 GTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCA--------------- 1995

Qy    659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
Db   1996 -----------------ACTGTTGAAGAAATTGTAGCTCCAAGTGTTGTAGAAAGTGTG 2037

Qy    679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
Db   2038 GCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTA------------ 2085

Qy    699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
Db   2086 ---------------GCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGT 2130

Qy    719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
Db   2131 GTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACT 2190

Qy    739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPro--------- 755
Db   2191 GTTGAAGAAATCGTAGCT------CCAACTGTTGAAGAAATTGTAGCTCCAAGTGTTGTA 2244

Qy    756 ---------ArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerVal 772
Db   2245 GAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCT 2304

Qy    773 GluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLys 792
Db   2305 GAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGT------------------ 2346

Qy    793 PheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAla------Thr 810
Db   2347 -----------------------GTAGCTGAAAATGTTGAAGAAAGTGTAGCTCCAACT 2382

Qy    811 ValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValVal 830
Db   2383 GTTGAAGAAATTGTAGCTCCAAGTGTTGAAGAA-----------------AGTGTAGCT 2424

Qy    831 ProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGlu 850
Db   2425 CCAAGTGTTGAAGAAAGTGTTGCTGAAAACGTTGCAACAAATTTATCAGACAATCTT--- 2481

Qy    851 LeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHis 870
Db   2482 TTAAGTAATTTATTAGGTGGTATC---GAAACTGAGGAAATAAAGGACAGTATATTAAAT 2538

Qy    871 ThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSer 890
Db   2539 GAGATAGAA--------------------------------GAAGTAAAAGAAAAT--- 2562

Qy    891 ProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMet 910

Db   2563 ---------GTAGTCACCACAATACTAGAAAACGTAGAAGAAACTACAGCTGAAAGTGTA 2613

Qy    911 ProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProValLeuThr 930
Db   2614 ACTACTTTTAGTAACATATTAGAGGAG---------ATACAAGAAAATACTATTACTAAT 2664

Qy    931 GlnThrLeuProGluGly---LysAspAlaHisAspAspMetValThrSerGluValAsp 949
Db   2665 GATACTATAGAGGAAAAATTAGAAGAACTCCACGAAAATGTATTAAGTGCCGCTTTAGAA 2724

Qy    950 PheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluVal 969
Db   2725 AATACCCAAAGTGAAGAGGAAAAGAAAGAAGTAATAGATGTAATT------GAAGAAGTA 2778

Qy    970 ThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGln------ 987
Db   2779 AAAGAA------------GAGGTCGCTACCACTTTAATAGAAACTGTGGAACAGGCAGAA 2826

Qy    988 -----------------LeuThrAspSerProAspThrThrGluGluAlaThrProVal 1001
Db   2827 GAAAAGAGCGCAAATACAATTACGGAAATATTTGAAAATTTAGAAGAA------------ 2874

Qy   1002 GlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIle 1021
Db   2875 AATGCAGTAGAAAGTAATGAAAATGTTGCAGAGAATTTAGAGAAATTAAACGAAACTGTA 2934

Qy   1022 LeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrVal 1041
Db   2935 TTTAATACTGTATTAGATAAAGTAGAGGAAACAGTAGAAATTAGCGGAGAAAGTTTAGAA 2994

Qy   1042 GlnArgThrGlySerLysAla---------------LeuGluLysValGluGluValGlu 1056
Db   2995 AACAATGAAATGGATAAAGCATTTTTTAGTGAAATATTTGATAATGTAAAAGGAATACAA 3054

Qy   1057 GluAsp--------------------------------SerGluValLeuAlaSerGlu 1065
Db   3055 GAAAATTTATTAACAGGTATGTTTCGAAGTATAGAAACCAGTATAGTAATCCAATCAGAA 3114

Qy   1066 LysGluLysAspVal-----------------------------------------Met 1071
Db   3115 GAAAAGGTTGATTTGAATGAAAATGTGGTTAGTTCGATTTTAGATAATATAGAAAATATG 3174

Qy   1072 ProLysGly-----------------------------------ProValGlnGluAla 1079
Db   3175 AAAGAAGGTTTATTAAATAAATTAGAAAATATTTCAAGTACTGAAGGTGTTCAAGAAACT 3234

Qy   1080 GlyAlaGluHisLeuAlaGlnGly------------------------------------ 1087
Db   3235 GTAACTGAACATGTAGAACAAAATGTATATGTGGATGTTGATGTTCCTGCTATGAAAGAT 3294

Qy   1088 -----------------SerGluThrGlyGlnAlaThrProGluSerLeuGluValPro 1101
Db   3295 CAATTTTTAGGAATATTAAATGAGGCAGGAGGGTTGAAAGAAATGTTTTTTAATTTGGAA 3354

Qy   1102 GluValThrAlaAspValAspHisValAlaThrCysGlnValIleLys---------Leu 1118
Db   3355 GATGTATTTAAAAGTGAAAGTGATGTAATTACTGTAGAAGAAATTAAGGATGAACCGGTT 3414

Qy   1119 GlnGlnLeuMetGluGln---------AlaValAlaProGluSerSerGluThrLeuThr 1135
Db   3415 CAAAAAGAGGTAGAAAAAGAAACTGTTAGTATTATTGAAGAAATGGAAGAAAATATTGTA 3474

Qy   1136 AspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGln 1155
Db   3475 GATGTATTAGAGGAAGAAAAAGAAGATTTAACAGACAAG---ATGATAGATGCAGTAGAA 3531

Qy   1156 GlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGln 1175
Db   3532 GAATCCATAGAAATATCTTCAGATTCTAAAGAAGAAACTGAATCTATTAAAGATAAA--- 3588

Qy   1176 ValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsn 1195
Db   3589 ------GAAAAAGATGTTTCACTAGTTGTTGAAGAA------------------------ 3618
```

```
Qy   1196 ValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGlu------Pro 1213
                      |||:::          :::|||       ||||||||
Db   3619 ------GTTCAAGACAATGATATGGATGAAAGTGTTGAGAAAGTTTTAGAATTGAAAAAT 3672

Qy   1214 ThrGlnGlnGluLeuThrAlaAlaAlaValProVal--------------------Leu 1226
             ::::::||||||      ||||||   :::                     :::
Db   3673 ATGGAAGAGGAGTTAATGAAGGATGCTGTTGAAATAAATGACATTACTAGCAAACTTATT 3732

Qy   1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu---AspGlyGluLysVal 1245
             :::|||:::           |||   |||   |||   :::   |||   ||||||:::
Db   3733 GAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATAAAAGATATGGAAAAATTA 3792

Qy   1246 LysGlu-----------------------GluGlnGluValPheValHisSerGlyPro 1257
          ||||||                           :::|||:::
Db   3793 AAAGAATTAGAAAAAGCATTATCAGAAGATTCTAAAGAAATAATAGATGCAAAAGATGAT 3852

Qy   1258 AsnSerGlnLysAlaAlaAspValThrTyrAsp---------------SerGluValMet 1272
                :::|||        :::      :::|||                  ||||||:::
Db   3853 ACATTAGAAAAAGTTATTGAAGAGGAACATGATATAACGACGACGTTGGATGAAGTTGTA 3912

Qy   1273 GlyValAlaGlyCysGlnGlu--------------------------------------- 1279
                :::      :::|||
Db   3913 GAATTAAAAGATGTCGAAGAAGACAAGATCGAAAAAGTATCTGATTTAAAAGATCTTGAA 3972

Qy   1280 ---------------------LysGluSerThrGluValGlnSerLeuSerLeuGluGlu 1292
                               ||||||      |||::::::|||      ||||||:::
Db   3973 GAAGATATATTAAAAGAAGTAAAAGAAATCAAAGAACTTGAAAGTGAAATTTTAGAAGAT 4032

Qy   1293 GlyGlu-----------MetGluThrAspValGluLysGluLysArgGluThrLysPro 1308
             :::               :::|||||||||:::   :::||||||:::|||   :::
Db   4033 TATAAAGAATTAAAAACTATTGAAACAGATATTTTAGAAGAGAAAAAAGAAATAGAAAAA 4092

Qy   1309 GluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyr 1328
          :::         :::      |||:::                 :::   |||
Db   4093 GATCATTTTGAAAAAATTCGAAGAAGAAGCTGAAGAAATAAAAGATCTTGAAGCAGATATA 4152

Qy   1329 GlyLysProValLeuThrLeuAspMetProSerSerGlu-------------------- 1341
             |||   |||   :::|||::::::            :::
Db   4153 TTAAAAGAAGTATCTTCATTAGAAGTTGAAGAAGAAAAAAAATTAGAAGAAGTACACGAA 4212

Qy   1341 ------------------------------------------------------------ 1341

Db   4213 TTAAAAGAAGAGGTAGAACATATAATAAGTGGTGATGCGCATATAAAAGGTTTGGAAGAA 4272

Qy   1342 --------------------------ArgGlyLysAlaLeuGlySerLeuGlyGlySer 1352
                                    :::|||      |||      |||   |||
Db   4273 GATGATTTAGAAGAAGTAGATGATTTAAAAGGAAGTATATTAGACATGTTAAAGGGAGAT 4332

Qy   1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
                |||   |||   ||||||                        :::||||||:::
Db   4333 ATGGAATTAGGGGATATGGATAAG-----------------------GAAAGTTTAGAA 4368

Qy   1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrValValIleSer 1392
                |||:::      |||   ||||||      ::::::   |||:::   |||
Db   4369 GATGTAACAACAAAACTTGGAGAAAGAGTTGAATCC---TTAAAAGATGTTTTATCTAGT 4425

Qy   1393 GluThrGlyGluSerProGluCysValGlyAlaHis----------------------- 1404
                |||         |||   :::
Db   4426 GCATTAGGCATGGATGAAGAACAAATGAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTG 4485

Qy   1405 ---------LeuLeu--------------------ProAlaGluLysSerSerAlaThr 1414
                   ||||||                    |||   :::|||   :::
Db   4486 GAAGAAGTATTATTAAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAATAACAAAAAAG 4545

Qy   1415 GlyGlyHisTrpThrLeuGlnHisAlaGlu------AspThrValProLeuGlyProGlu 1432
                :::   ::::::      |||      |||   :::   :::      |||
Db   4546 AAAGTAAGGTTTGATATTAAGGATAAGGAACCAAAAGATGAAATA---GTAGAAGTTGAA 4602

Qy   1433 SerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisPro 1452
             :::   |||   |||                           |||      :::
Db   4603 ATGAAAGATGAAGATATA-----------------------GAAGAAGATGTAGAAGAA 4638

Qy   1453 AspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysPro 1472
             |||::::::   :::|||            :::   |||      :::|||:::
Db   4639 GATATAGAAGAAGATATAGAAGAAGATAAAGTTGAAGATATAGATGAAGATATAGATGAA 4698

Qy   1473 AspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeuLysAla 1492
             |||   |||   |||   |||                     |||:::   :::::::::
Db   4699 GATATAGGTGAAGACAAAGAT-----------GAAGTTATAGATTTAATAGTCCAAAAA 4746

Qy   1493 GluProGluIleLeuGluLeuGluSerLysSerAsnLysIle 1506
             |||      |||   ::::::::::::|||      |||:::
Db   4747 GAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAAAATTA 4788
```

RESULT 7
US-09-742-096-1
; Sequence 1, Application US/09742096
; Patent No. US20020155441A1
; GENERAL INFORMATION:
; APPLICANT: DRUILHE, PIERRE
; APPLICANT: DAUBERSIES, PIERRE
; TITLE OF INVENTION: MALARIAL PRE-ERYTHROCYTIC STAGE POLYPEPTIDE MOLECULES
; FILE REFERENCE: 200773US0DIV
; CURRENT APPLICATION NUMBER: US/09/742,096
; CURRENT FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 08/973,642
; PRIOR FILING DATE: 1998-02-06
; PRIOR APPLICATION NUMBER: PCT/FR96/00894
; PRIOR FILING DATE: 1996-06-12
; PRIOR APPLICATION NUMBER: FR 95/07007
; PRIOR FILING DATE: 1995-06-13
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
;   LENGTH: 6152
;   TYPE: DNA
;   ORGANISM: P. falciparum
US-09-742-096-1

```
Alignment Scores:
Pred. No.:               1.97e-23     Length:        6152
Score:                   462.00       Matches:       323
Percent Similarity:      36.36%       Conservative:  293
Best Local Similarity:   19.07%       Mismatches:    646
Query Match:             5.72%        Indels:        432
DB:                      9            Gaps:          68
```

US-09-902-432-4 (1-1596) x US-09-742-096-1 (1-6152)

```
Qy     53 GlnLysAsnGlyGlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisVal 72
          |||   ||||||:::|||:::|||                  |||:::
Db    527 CAAGTAAATGGTGAATTAGCTAGT---------------GAAGAAGTAAAGGAAAAAATT 571

Qy     73 GlnGluGluAsnGlnGluGlyGlnGlu---GluGluValValAspGluAspValGlyGln 91
             :::      :::||||||            |||   ||||||::::::
Db    572 CTTGACTTATTAGAAGAAGGAAATACATTAACTGAAAGTGTAGATGATAATAAAAATTTA 631

Qy     92 ArgGluSerGluAspValArgGluLys---------AspArgValGluGlu--------- 105
             |||:::||||||::::::|||                  :::||||||
Db    632 GAAGAAGCCGAAGATATAAAGGAAAATATCTTATTAAGTAATATAGAAGAACCAAAAGAA 691

Qy    106 --------------------------MetAlaAlaAsnSerThrAlaValGluAspIle 116
                                    :::      ||||||      |||   :::
Db    692 AATATTATTGACAATTTATTAAATAATATTGGACAAAATTCAGAAAAACAAGAAAGTGTA 751

Qy    117 ThrLysAspGlyGln---------GluGluThrSerGluIleIleGluGlnIleProAla 133
          ::::::::   |||            |||      :::|||:::::        :::
Db    752 TCAGAAAATGTACAAGTCAGTGATGAACTTTTTAATGAATTATTAAATAGTGTAGATGTT 811

Qy    134 SerGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGly 153
          :::               |||   :::::::::      |||||||||   ||||||
Db    812 AATGGAGAAGTAAAAGAAAATATTTTGGAG------GAAAGTCAAGTTAATGACGATATT 865

Qy    154 PheLysLysValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGlu 173
          |||      :::   |||                        :::|||::::::::::::
```

```
Db    866 TTTAATAGTTTAGTAAAA-----------------AGTGTTCAACAAGAACAACAACAC 907

Qy    174 LysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAla 193
Db    908 AATGTTGAAGAAAAAGTTGAAGAAAGTGTAGAAGAAAATGACGAAGAAAGTGTAGAAGAA 967

Qy    194 SerVal------GlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGlu 211
Db    968 AATGTAGAAGAAAATGTAGAAGAAAATGACGACGGAAGTGTAGCCTCAAGTGTTGAAGAA 1027

Qy    212 Ser------AlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThr 229
Db   1028 AGTATAGCTTCAAGTGTTGATGAAAGTATAGATTCAAGTATTGAAGAAAATGTAGCTCCA 1087

Qy    230 LeuLysGlnGlu-------------------------------------------------- 233
Db   1088 ACTGTTGAAGAAATCGTAGCTCCAAGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAA 1147

Qy    234 ------GlnSerSerThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGlu 251
Db   1148 AGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAA 1207

Qy    252 GluAlaLysAspGluGlyGluGluLysGlnGluLysGluProThrLysSerProGluSer 271
Db   1208 AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAA 1267

Qy    272 ProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGly 291
Db   1268 ATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTGTAGCTCCA 1327

Qy    292 TrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGlu--- 310
Db   1328 AGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAA 1387

Qy    311 ---ThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGlu 329
Db   1388 AGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAA 1447

Qy    330 LysThrGluProAla---------SerGluGluGlnGluProAlaGluAspThrAspGln 346
Db   1448 AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAA 1507

Qy    347 AlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAsp 366
Db   1508 ATC---GTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATT------ 1558

Qy    367 LeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGluValPheAspGluLys 386
Db   1559 GTAGCTCCAAGTGTTGTAGAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAAT 1618

Qy    387 MetGluAlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGlu 406
Db   1619 GTTGAA------GAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAA 1672

Qy    407 GluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGlu 426
Db   1673 GAAAGT-----------------------------GTAGCTGAAAATGTTGAAGAA 1699

Qy    427 SerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGluProAla 446
Db   1700 AGTGTA------------------------------------------GCTGAAAATGTT 1717

Qy    447 GluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThr 466
Db   1718 GAAGAAAGTGTAGCTGAA------------------------------------------ 1735

Qy    467 AspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluVal 486
Db   1736 ------AATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTT 1789

Qy    487 GluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLysLeuPhe 506
Db   1790 GAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATTGTA 1849

Qy    507 SerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGly 526
Db   1850 GCTCCAAGTGTTGTAGAAAGTGTGGCT--------------------------------- 1876

Qy    527 GlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAsp 546
Db   1877 ------------CCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCT 1924

Qy    547 GluGlnLysGlyGluSerSerAlaSerSerProGluGluPro---------------Glu 561
Db   1925 GAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAA 1984

Qy    562 GluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGlu 581
Db   1985 GAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCT 2044

Qy    582 GlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLys 601
Db   2045 CCAACTGTT-----------------GAAGAAATTGTAGCTCCA------AGTGTTGTA 2080

Qy    602 LysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGlu 621
Db   2081 GAAAGTGTGGCTCCAAGT------GTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGT 2134

Qy    622 Leu---GluLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMet 640
Db   2135 GTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAAT 2194

Qy    641 GlnAspGluVal------LysThrValGlyGluGluGlnLysProGluGluProLysArg 658
Db   2195 GTTGAAGAAATCGTAGCTCCAACTGTTGAAGAAATCGTAGCTCCA--------------- 2239

Qy    659 ArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArg 678
Db   2240 -----------------ACTGTTGAAGAAATTGTAGCTCCAAGTGTTGTAGAAAGTGTG 2281

Qy    679 AlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSer 698
Db   2282 GCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTA------------ 2329

Qy    699 HisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSer 718
Db   2330 ---------------GCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGT 2374

Qy    719 ThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSer 738
Db   2375 GTAGCTGAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAATCGTAGCTCCAACT 2434

Qy    739 GluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPro--------- 755
Db   2435 GTTGAAGAAATCGTAGCT------CCAACTGTTGAAGAAATTGTAGCTCCAAGTGTTGTA 2488

Qy    756 ---------ArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerVal 772
Db   2489 GAAAGTGTGGCTCCAAGTGTTGAAGAAAGTGTAGAAGAAAATGTTGAAGAAAGTGTAGCT 2548

Qy    773 GluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLys 792
Db   2549 GAAAATGTTGAAGAAAGTGTAGCTGAAAATGTTGAAGAAAGT------------------ 2590

Qy    793 PheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAla------Thr 810
Db   2591 -----------------------GTAGCTGAAAATGTTGAAGAAAGTGTAGCTCCAACT 2626

Qy    811 ValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValVal 830
Db   2627 GTTGAAGAAATTGTAGCTCCAAGTGTTGAAGAA------------------AGTGTAGCT 2668

Qy    831 ProLeuSerGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGlu 850
Db   2669 CCAAGTGTTGAAGAAAGTGTTGCTGAAAACGTTGCAACAAATTTATCAGACAATCTT--- 2725
```

```
Qy     851 LeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHis 870
           |||      ||||||||    :::     ::||||||:::      :::::::::::
Db    2726 TTAAGTAATTTATTAGGTGGTATC---GAAACTGAGGAAATAAAGGACAGTATATTAAAT 2782

Qy     871 ThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSer 890
           :::                                            |||:::|||
Db    2783 GAGATAGAA--------------------------------GAAGTAAAAGAAAAT--- 2806

Qy     891 ProSerTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMet 910
                    :::      ::::::  |||   :::|||  |||      |||:::::
Db    2807 ---------GTAGTCACCACAATACTAGAAAACGTAGAAGAAACTACAGCTGAAAGTGTA 2857

Qy     911 ProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProValLeuThr 930
                          :::   |||:::               :::|||     :::
Db    2858 ACTACTTTTAGTAACATATTAGAGGAG---------ATACAAGAAAATACTATTACTAAT 2908

Qy     931 GlnThrLeuProGluGly---LysAspAlaHisAspAspMetValThrSerGluValAsp 949
              |||:::   |||      ::::::  |||::::::::::::::::::   ::::::
Db    2909 GATACTATAGAGGAAAAATTAGAAGAACTCCACGAAAATGTATTAAGTGCCGCTTTAGAA 2968

Qy     950 PheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluVal 969
              |||                     |||    :::   :::     ||||||||
Db    2969 AATACCCAAAGTGAAGAGGAAAAGAAAGAAGTAATAGATGTAATT------GAAGAAGTA 3022

Qy     970 ThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGln------ 987
           |||               |||    |||  ::::::       |||   |||
Db    3023 AAAGAA------------GAGGTCGCTACCACTTTAATAGAAACTGTGGAACAGGCAGAA 3070

Qy     988 ------------------LeuThrAspSerProAspThrThrGluGluAlaThrProVal 1001
                             :::|||:::     :::     |||||
Db    3071 GAAAAGAGCGCAAATACAATTACGGAAATATTTGAAAATTTAGAAGAA------------ 3118

Qy    1002 GlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIle 1021
                 ||||||||                |||      |||:::       :::   :::
Db    3119 AATGCAGTAGAAAGTAATGAAAATGTTGCAGAGAATTTAGAGAAATTAAACGAAACTGTA 3178

Qy    1022 LeuGlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrVal 1041
                    |||   ||||||||:::|||        ::::::
Db    3179 TTTAATACTGTATTAGATAAAGTAGAGGAAACAGTAGAAATTAGCGGAGAAAGTTTAGAA 3238

Qy    1042 GlnArgThrGlySerLysAla---------------LeuGluLysValGluGluValGlu 1056
                          ||||||               :::   |||:::     ::::::
Db    3239 AACAATGAAATGGATAAAGCATTTTTTAGTGAAATATTTGATAATGTAAAAGGAATACAA 3298

Qy    1057 GluAsp---------------------------------SerGluValLeuAlaSerGlu 1065
           |||:::                                  |||   |||:::  ||||||
Db    3299 GAAAATTTATTAACAGGTATGTTTCGAAGTATAGAAACCAGTATAGTAATCCAATCAGAA 3358

Qy    1066 LysGluLysAspVal------------------------------------------Met 1071
           ::::::   |||:::                                           |||
Db    3359 GAAAAGGTTGATTTGAATGAAAATGTGGTTAGTTCGATTTTAGATAATATAGAAAATATG 3418

Qy    1072 ProLysGly------------------------------------ProValGlnGluAla 1079
              :::|||                                     |||||||||
Db    3419 AAAGAAGGTTTATTAAATAAATTAGAAAATATTTCAAGTACTGAAGGTGTTCAAGAAACT 3478

Qy    1080 GlyAlaGluHisLeuAlaGlnGly------------------------------------ 1087
                 ||||||:::   |||
Db    3479 GTAACTGAACATGTAGAACAAAATGTATATGTGGATGTTGATGTTCCTGCTATGAAAGAT 3538

Qy    1088 ------------------SerGluThrGlyGlnAlaThrProGluSerLeuGluValPro 1101
                             :::|||   |||                        :::
Db    3539 CAATTTTTAGGAATATTAAATGAGGCAGGAGGGTTGAAAGAAATGTTTTTTAATTTGGAA 3598

Qy    1102 GluValThrAlaAspValAspHisValAlaThrCysGlnValIleLys---------Leu 1118
           :::|||                   |||   |||   :::   ||||||         :::
Db    3599 GATGTATTTAAAAGTGAAAGTGATGTAATTACTGTAGAAGAAATTAAGGATGAACCGGTT 3658

Qy    1119 GlnGlnLeuMetGluGln---------AlaValAlaProGluSerSerGluThrLeuThr 1135
           |||:::   :::|||:::          ::::::      |||       |||   :::
Db    3659 CAAAAAGAGGTAGAAAAAGAAACTGTTAGTATTATTGAAGAAATGGAAGAAAATATTGTA 3718

Qy    1136 AspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGln 1155
           |||                           |||   |||            |||   :::
Db    3719 GATGTATTAGAGGAAGAAAAAGAAGATTTAACAGACAAG---ATGATAGATGCAGTAGAA 3775

Qy    1156 GlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSerGln 1175
           :::         |||   |||         :::   |||   ::::::::
Db    3776 GAATCCATAGAAATATCTTCAGATTCTAAAGAAGAAACTGAATCTATTAAAGATAAA--- 3832

Qy    1176 ValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsn 1195
                 |||::::::   :::         :::|||
Db    3833 ------GAAAAAGATGTTTCACTAGTTGTTGAAGAA------------------------ 3862

Qy    1196 ValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGlu------Pro 1213
                 |||:::         :::|||         |||||||||
Db    3863 ------GTTCAAGACAATGATATGGATGAAAGTGTTGAGAAAGTTTTAGAATTGAAAAAT 3916

Qy    1214 ThrGlnGlnGluLeuThrAlaAlaAlaValProVal---------------------Leu 1226
              ::::::||||||         ||||||   :::                     :::
Db    3917 ATGGAAGAGGAGTTAATGAAGGATGCTGTTGAAATAAATGACATTACTAGCAAACTTATT 3976

Qy    1227 AlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeu---AspGlyGluLysVal 1245
              :::|||:::         |||   |||   |||   :::   |||   ||||||:::
Db    3977 GAAGAAACTCAAGAGTTAAATGAAGTAGAAGCAGATTTAATAAAAGATATGGAAAAATTA 4036

Qy    1246 LysGlu-----------------------GluGlnGluValPheValHisSerGlyPro 1257
           ||||||                                 :::|||:::
Db    4037 AAAGAATTAGAAAAAGCATTATCAGAAGATTCTAAAGAAATAATAGATGCAAAAGATGAT 4096

Qy    1258 AsnSerGlnLysAlaAlaAspValThrTyrAsp---------------SerGluValMet 1272
                 :::|||       :::      :::|||               ||||||:::
Db    4097 ACATTAGAAAAAGTTATTGAAGAGGAACATGATATAACGACGACGTTGGATGAAGTTGTA 4156

Qy    1273 GlyValAlaGlyCysGlnGlu--------------------------------------- 1279
              :::      :::|||
Db    4157 GAATTAAAAGATGTCGAAGAAGACAAGATCGAAAAAGTATCTGATTTAAAAGATCTTGAA 4216

Qy    1280 --------------------LysGluSerThrGluValGlnSerLeuSerLeuGluGlu 1292
                               ||||||   ||||::::::|||      ||||||:::
Db    4217 GAAGATATATTAAAAGAAGTAAAAGAAATCAAAGAACTTGAAAGTGAAATTTTAGAAGAT 4276

Qy    1293 GlyGlu------------MetGluThrAspValGluLysGluLysArgGluThrLysPro 1308
              :::            :::|||||||||:::   :::||||||:::|||   :::
Db    4277 TATAAAGAATTAAAAACTATTGAAACAGATATTTAGAAGAGAAAAAAGAAATAGAAAAA 4336

Qy    1309 GluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyr 1328
           :::         :::      |||:::            :::      |||
Db    4337 GATCATTTTGAAAAATTCGAAGAAGAAGCTGAAGAAATAAAAGATCTTGAAGCAGATATA 4396

Qy    1329 GlyLysProValLeuThrLeuAspMetProSerSerGlu--------------------- 1341
              |||   |||   :::|||::::::         :::
Db    4397 TTAAAAGAAGTATCTTCATTAGAAGTTGAAGAAGAAAAAAATTAGAAGAAGTACACGAA 4456

Qy    1341 ------------------------------------------------------------ 1341

Db    4457 TTAAAAGAAGAGGTAGAACATATAATAAGTGGTGATGCGCATATAAAAGGTTTGGAAGAA 4516

Qy    1342 --------------------------ArgGlyLysAlaLeuGlySerLeuGlyGlySer 1352
                                     :::|||      |||      |||   |||
Db    4517 GATGATTTAGAAGAAGTAGATGATTTAAAAGGAAGTATATTAGACATGTTAAAGGGAGAT 4576

Qy    1353 ProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAsp 1372
              |||   |||   ||||||                        :::||||||:::
Db    4577 ATGGAATTAGGGGATATGGATAAG------------------------GAAAGTTTAGAA 4612

Qy    1373 ThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrValValIleSer 1392
                    |||:::      |||   ||||||      ::::::   |||:::   |||
Db    4613 GATGTAACAACAAAACTTGGAGAAAGAGTTGAATCC---TTAAAAGATGTTTTATCTAGT 4669

Qy    1393 GluThrGlyGluSerProGluCysValGlyAlaHis------------------------ 1404
              |||         |||   :::
Db    4670 GCATTAGGCATGGATGAAGAACAAATGAAAACAAGAAAAAAAGCTCAAAGACCTAAGTTG 4729

Qy    1405 ---------LeuLeu--------------------ProAlaGluLysSerSerAlaThr 1414
                    ||||||                    |||   :::|||   :::
```

```
Db    4730 GAAGAAGTATTATTAAAAGAAGAGGTTAAAGAAGAACCAAAGAAAAAAATAACAAAAAAG 4789

Qy    1415 GlyGlyHisTrpThrLeuGlnHisAlaGlu------AspThrValProLeuGlyProGlu 1432
                    :::   ::::::      |||       |||   :::   :::      |||
Db    4790 AAAGTAAGGTTTGATATTAAGGATAAGGAACCAAAAGATGAAATA---GTAGAAGTTGAA 4846

Qy    1433 SerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHisPro 1452
                 :::   |||   |||                        |||      :::
Db    4847 ATGAAAGATGAAGATATA-----------------------GAAGAAGATGTAGAAGAA 4882

Qy    1453 AspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLysPro 1472
           |||::::::   :::|||            :::   |||      :::|||:::
Db    4883 GATATAGAAGAAGATATAGAAGAAGATAAAGTTGAAGATATAGATGAAGATATAGATGAA 4942

Qy    1473 AspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeuLysAla 1492
           |||   |||   |||   |||                     |||:::   :::::::::
Db    4943 GATATAGGTGAAGACAAAGAT------------GAAGTTATAGATTTAATAGTCCAAAAA 4990

Qy    1493 GluProGluIleLeuGluLeuGluSerLysSerAsnLysIle 1506
           |||      |||   ::::::::::::|||      |||:::
Db    4991 GAGAAACGCATTGAAAAGGTTAAAGCGAAAAAGAAAAAATTA 5032
```

RESULT 8
US-09-864-761-17726
; Sequence 17726, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 17726
;   LENGTH: 7997
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO AL034555.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 10
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 8.9
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 4.8
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 14
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 7.2
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.5
;   OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.1
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 9.3
;   OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 7.7
;   OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 12
;   OTHER INFORMATION: EST_HUMAN HIT: AU117052.1, EVALUE 0.00e+00
;   OTHER INFORMATION: SWISSPROT HIT: P08640, EVALUE 4.00e-10
;   OTHER INFORMATION: NT HIT: gi11427711, EVALUE 0.00e+00
US-09-864-761-17726

Alignment Scores:

Pred. No.:	3.69e-21	Length:	7997
Score:	431.50	Matches:	373
Percent Similarity:	34.27%	Conservative:	271
Best Local Similarity:	19.85%	Mismatches:	759
Query Match:	5.34%	Indels:	476
DB:	10	Gaps:	83

US-09-902-432-4 (1-1596) x US-09-864-761-17726 (1-7997)

```
Qy      31 ProAlaAlaGluAlaSerGlyAlaAlaGly-AspProAlaAspAlaAspProAla----- 48
           |||      |||   ||||||   :::      :::||||||    :::|||   :::
Db     886 CCTGGACCAGAAACTTCAGGTCTCTCAGACGGAGCCTGCAAAATCTGACTTGTCTAAACT 945

Qy      49 ------------ThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAsnGlyVa 64
                       |||:::|||:::::      |||       |||    |||       ||
Db     946 GGAATCAGTTAGAATGAAAGTACCAAAGGAAAAGGGGCTTTCAAGCCATGTTGAAGTGGT 1005

Qy      64 lAlaGluGlnGlyAspVal-----------HisValGlnGluGluAsnGlnGluGlyGl 80
           |    ::::::|||   :::           |||::::::   |||   |||
Db    1006 GGAGAAGGAAGGCAGGCTTAAAGCCAGGAAGCACCTCAAGCCTGAG---CAGCCTGCAGA 1062

Qy      80 nGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLy 100
                          ||||||
Db    1063 TGGGGTAAGTGCTGTGGAT----------------------------------------- 1081

Qy     100 sAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGl 120
                  :::|||::::::   |||                     |||
Db    1082 -------CTGGAGAAGCTGGAAGCCAGGAAAAGGCGCTTTGCAGATTCCAATTTAAAAGC 1134

Qy     120 yGlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsn---AsnValGl 139
            :::::::::         |||:::   :::      |||      |||:::      |||
Db    1135 AGAAAAGCAAAAACCAGAGGTCAAGAAAAGCAGTCCAGAGATGGAGGATGCTCGCGTGCT 1194

Qy     139 uGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPhe-- 158
                          ||||||                  ||||||      ::::::|||
Db    1195 TTCAAAAAAGCAGCCT------------------GACGTGTCCTCTAGAGAGGTCATTCT 1236

Qy     159 ------------------------LysPheValGlyPheLysPheThrValLysLysAs 170
                                   |||   |||         :::         ::::::
Db    1237 GCTGAGGGAAGGAGAGGCTGAAAGAAAGCCTGTGAGGAAAGAAATTCTTAAAAGAGAATC 1296

Qy     170 pLysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGl 190
             |||      |||   |||   :::         :::         ||||||      |||
Db    1297 TAAAAAAATCAAACTGGACAGACTTAATACTGTTGCCAGCCCCAAAGACTGTCAGGAGCT 1356

Qy     190 yAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGl 210
             |||      |||||||||:::|||                    ::::::      :::||
Db    1357 TGCCAGTATTTCTGTTGGGTCTGGCTCAAGGCCCAGCTCAGACCTACAAGCAAGACTGGG 1416

Qy     210 yGluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLe 230
```

```
Db   1417 AGAACTAGCAGGT------------------GAATCTGTGGAAAATCAAGAA-------- 1450

Qy    230 uLysGlnGluGlnSerSerThrGluIlePro--------------------LeuGlnAl 243

Db   1451 -------GTCCAATCAAAAAAGCCCATTCCCTCAAAACCACAGCTCAAACAGCTGCAGGT 1503

Qy    243 aGluSerAspGlnAlaAlaGluGluGlu------AlaLysAspGluGlyGluGluLysGl 261

Db   1504 ATTAGATGATCAAGGACCAGAGAGAGAAGACGTTAGGAAAAACTATTGCAGTCTTCGTGA 1563

Qy    261 nGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThr-- 280

Db   1564 TGAAACACCTGAACGTAAATCAGGCCAAGAGAAATCACATTCAGTAAATACTGAAGAAAA 1623

Qy    281 ---------------------ThrSerSerPheLysLysPhePheThrHisGlyTrpAl 293

Db   1624 AATTGGCATTGACATCGATCACACGCAGAGTTACCGAAAACAAATGGAACAGAGTCGTAG 1683

Qy    293 aGlyTrpArgLysLysThrSerPheLysLysSerLys---------------GluAspAs 308

Db   1684 GAAACAGCAGATGGAAATGGAAATAGCCAAGTCTGAGAAGTTTGGCAGTCCTAAAAAAGA 1743

Qy    308 pLeuGluThrAlaGluLysArg------------------------------------- 315

Db   1744 TGTAGATGAATATGAAAGACGTAGCCTCGTTCACGAGGTAGGCAAACCCCCTCAAGATGT 1803

Qy    316 ---------------------LysGluGlnGluAlaGluLysValAsp----------- 324

Db   1804 CACTGATGACTCTCCTCCTAGCAAAAAGAAAAGGATGGATCATGTCGATTTTGATATCTG 1863

Qy    325 ----------GluGluGluLysGluLysThrGluProAlaSerGluGluGlnGluProAl 341

Db   1864 CACCAAGCGAGAACGGAATTACAGAAGTTCACGCCAAATCAGCGAAGATTCTGAAAGGAC 1923

Qy    341 aGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGl 361

Db   1924 TGGTGGTTCTCCCAGTGTCCGACATGGTTCCTTCCATGAAGAT---------------GA 1968

Qy    361 uAspGlnValGlyAspLeuGluAlaSerSer--------------------GluGluLy 374

Db   1969 GGATCCCATAGGCTCCCCTAGGCTACTGTCAGTAAAAGGGTCTCCTAAAGTAGATGAAAA 2028

Qy    374 sCysAlaPro-----------------------------LeuAlaThrGluValPheAs 384

Db   2029 AGTCCTCCCCTATTCTAACATAACAGTCAGGGAAGAGTCTTTAAAATTTAATCCTTATGA 2088

Qy    384 pGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSerThrValGluLy 404

Db   2089 TTCTAGCAGGAGAGAACAGATGGCAGATATGGCCAAAATAAAACTATCTGTCTTGAATTC 2148

Qy    404 sThrGluGluGluGlnGlyGlyGlyGlyGluAlaGlu---------GlyGlyValValVa 421

Db   2149 TGAAGATGAACTAAATCGTTGGGACTCTCAGATGAAACAGGATGCTGGCAGATTTGATGT 2208

Qy    421 lGluGlyThrGlyGluSerLeuProProGluLysLeu--------------------Al 434

Db   2209 GAGTTTCCCAAACAGCATAATTAAGAGAGATAGCCTTCGAAAAAGGTCTGTACGAGATCT 2268

Qy    434 aGluProGlnGluValProGlnGluAlaGluProAlaGluGlu---------------- 448

Db   2269 GGAACCTGGTGAGGTGCCTTCTGATTCTGACGAAGATGGTGAACACAAATCCCACTCACC 2328

Qy    449 ---------------------------------------LeuMetLysSerArgGluMe 455

Db   2329 CAGAGCCTCTGCATTATATGAAAGTTCTCGATTGTCTTTTTATTGAGGGACAGAGAAGA 2388

Qy    455 tCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGlu----------- 471

Db   2389 CAAGCTACGTGAGCGAGATGAAAGACTCTCTAGTTCTTTAGAAAGGAACAAATTTTACTC 2448

Qy    472 ----------GluLysThrLeuProLysHisProGluGlyIleValSerGluValGluMe 488

Db   2449 TTTTGCATTGGATAAGACAATCACACCAGACACTAAAGCTTTGCTTGAAAGAGCTAAATC 2508

Qy    488 tLeuSerSer-------------------------------------------------- 491

Db   2509 CCTCTCTTCATCTCGTGAAGAAAATTGGTCTTTTCTTGATTGGGACTCCCGATTTGCAAA 2568

Qy    492 -------------GlnGluArgIleLysValGlnGlySerPro---------------- 501

Db   2569 TTTTCGAAACAACAAAGATAAAGAAAAGGTTGACTCTGCTCCAAGACCTATTCCATCCTG 2628

Qy    502 -------LeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGl 519

Db   2629 GTACATGAAAAAGAAGAAAATTAGGACTGATTCAGAAGGGAAAATGGATGATAAGAAAGA 2688

Qy    519 nLysGlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisTh 539

Db   2689 GGACCATAAA---------------GAAGAAGAGCAAGAGAGGCAGGAATTGTTTGCTTC 2733

Qy    539 rGluSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGl 559

Db   2734 TCGTTTTTTACACAGCTCAATCTTTGAACAAGATTCCAAGCGATTGCAGCATCTAGAGAG 2793

Qy    559 uProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAl 579

Db   2794 AAAAGAGGAAGATTCTGACTTCATTTCTGGTAGGATCTATGGGAAGCAGACATCTGAGGG 2853

Qy    579 aGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSe 599

Db   2854 AGCAAACAGCACAACT---------GATTCCATTCAAGAACCAGTAGTTCTGTTCCATAG 2904

Qy    599 r---PheLysLysMetVal------------ThrProLysLysArgValArgArgProSe 614

Db   2905 CAGATTTATGGAGCTCACACGGATGCAACAGAAAGAAAAAGAAAAAGACCAGAAACCCAA 2964

Qy    614 rGluSerAspLysGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAs 634

Db   2965 AGAGGTTGAGAAACAGGAAGATACAGAGAATCATCCCAAGACCCCAGAATCTGCTCCTGA 3024

Qy    634 pSerThrValSerGluMetGln---------------------AspGluValLysThrVa 647

Db   3025 GAATAAAGATTCAGAACTGAAAACTCCACCTTCCGTTGGGCCTCCAAGTGTCACAGTCGT 3084

Qy    647 lGlyGluGluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGl 667

Db   3085 AACTCTAGAATCAGCCCCATCAGCACTAGAGAAGACCACTGGTGACAAAACGGTAGAGGC 3144

Qy    667 uAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAs 687

Db   3145 GCCTTTGGTA---------ACAGAAGAGAAGACTGTGGAGCCAGCTACCGTCTCAGAAGA 3195

Qy    687 p---------------GluGlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGl 702

Db   3196 AGCAAAGCCTGCATCTGAACCTGCTCCTGCCCCTGTG------------GAACAGCTGGA 3243

Qy    702 uGlu-----------------AlaSerLysAspLysGluAlaGlyThrAspAlaValPr 716

Db   3244 ACAAGTAGACCTGCCCCCAGGAGCAGACCCCGATAAAGAAGCTGCC------ATGATGCC 3297

Qy    716 oAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerProGluPro-------- 733

Db   3298 TGCGGGTGTTGAGGAA---------------GGTTCATCAGGTGACCAGCCGCCTTATCT 3342

Qy    734 ----AlaGlySerProSerGluGlyGluGlyValSerThrTrpGlu-------------- 747

Db   3343 GGATGCCAAGCCTCCAACTCCCGGGGCCTCGTTTTCCCAGGCAGAGAGCAACGTAGATCC 3402

Qy    748 ----------SerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGl 764

Db   3403 AGAGCCTGACAGTACCCAGCCACTTTCAAAACCAGCTCAGAAGTCTGAGGAAGCCAATGA 3462

Qy    764 uGluLysAlaGlu------AspSerSerValGlu------------------------- 773

Db   3463 GCCAAAGGCCGAAAAGCCAGACGCCACTGCAGATGCTGAGCCTGATGCAAACCAGAAAGC 3522
```

```
Qy    774 -GlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPh 793
Db   3523 CGAAGCTGCTCCTGAGTCTCAGCCCCCAGCTTCTGAAGAT---TTAGAGGTTGATCCTCC 3579

Qy    793 eIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAs 813
Db   3580 AGTTGCTGCAAAGGATAAAAAGCCAAACAAAAGCAAGCGTTCAAAGACCCCTGTTCAG-- 3637

Qy    813 pSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValValProLeuSe 833
Db   3638 -GCAGCTGCAGTGAGTATCGTGGAG---------------AAGCCCGTCACAAGGAAGAG 3681

Qy    833 rGluTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThr------GluLe 851
Db   3682 TGAG------AGGATAGACCGGGAAAAACTCAAGCGGTCCAATTCTCCTCGGGGAGAAGC 3735

Qy    851 uProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThr----------- 867
Db   3736 ACAGAAGCTTTTGGAATTGAAGATGGAGGCAGAGAAGATTACAAGGACTGCTTCTAAAAA 3795

Qy    868 -------------LeuValHisThrValSerValAlaValIleAspGlyThrArgAla-- 882
Db   3796 CTCTGCTGCAGACCTTGAACATCCCGAACCAAGTTTGCCTCTCAGCCGAACAAGGCGCCG 3855

Qy    883 ----ValThrSerVal---------------------GluGluArgSerProSerTrpIl 894
Db   3856 GAATGTAAGGAGCGTCTATGCAACCATGGGTGACCATGAAAACCGCTCTCCT-------- 3907

Qy    894 eSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGl 914
Db   3908 ----------GTCAAAGAGCCCGTTGAGCAA---------------------------CC 3930

Qy    914 uGluValThrGluLysAspIleIleAlaGlu------GluThrProValLeuThrGlnTh 932
Db   3931 AAGAGTGACCAGAAAGAGATTGGAGCGAGAGCTTCAGGAGGCTGCAGCGGTTCCCACCAC 3990

Qy    932 r---------------LeuProGluGlyLysAspAlaHisAspAspMetValThrSerGl 947
Db   3991 CCCTCGGAGGGGAAGGCCTCCAAAGACACGCCGGCGAGCCGATGAAGAGGAGGAGAACGA 4050

Qy    947 uValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGl 967
Db   4051 GGCCAAGGAACCTGCAGAAACACTCAAGCCACCTGAGGGATGGCGGTCGCCAAGGTCCCA 4110

Qy    967 uGluValThrGluAlaSerGlyAlaGlu------------GluThrThrAspMetValSe 983
Db   4111 GAAAACTGCAGCTGGTGGTGGACCCCAAGGGAAAAAGGGAAAAAATGAACCGAAGGTGGA 4170

Qy    983 rAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGl 1003
Db   4171 TGCTACA---------------CGTCCTGAGGCCACCACTGAGGTGGGCCCCCAAATAGG 4215

Qy   1003 uValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGl 1023
Db   4216 CGTGAAAGAGAGCTCCATGGAACCCAAGGCTGCTGAGGAGGAGGCAGGGAGTGAACAGAA 4275

Qy   1023 nAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnAr 1043
Db   4276 ACGTGACAGAAAAGATGCTGGCACAGACAAAAACCCCCCTGAAACCGCCCCTGTTGAA-- 4333

Qy   1043 gThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSer----------- 1059
Db   4334 -------------GTTGTAGAGAAAAAACCGGCCCCTGAAAAAAACTCCAAATCAAAGAG 4380

Qy   1060 -------------------------------GluValLeuAlaSerGluLysGluLysAs 1069
Db   4381 AGGAAGATCTCGAAACTCCAGGTTAGCAGTGGACAAATCTGCAAGTCTGAAAAATGTGGA 4440

Qy   1069 p------ValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGl 1087
Db   4441 TGCTGCTGTCAGTCCCAGGGGGGCTGCAGCACAGGCAGGGGAGAGG-------------- 4486

Qy   1087 ySerGluThrGly------GlnAlaThrProGluSerLeuGluValProGluValThrAl 1105
Db   4487 ----GAATCTGGGGTGGTGGCAGTCTCCCCTGAGAAAAGTGAGAGTCCCCAAAAGGAGGA 4542

Qy   1105 aAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnAl 1125
Db   4543 TGGTTTATCA------------------------------TCCCAGTTGAAAAGTGATCC 4572

Qy   1125 aValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLe 1145
Db   4573 AGTTGATCCAGACAAGGAACCAGAGAAAGAAGACGTGTCTGCCTCTGGGCCGTCCCCAGA 4632

Qy   1145 uAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSer-------- 1162
Db   4633 AGCC---ACCCAGTTAGCCAAGCAGATGGAGCTGGAGCAGGCCGTGGAACACATCGCAAA 4689

Qy   1163 ----GlnAspSerLysAlaThrAlaAlaValArgGlnSer-------------------- 1174
Db   4690 GCTCGCTGAGGCCTCTGCCTCTGCTGCCTATAAGGCAGATGCACCAGAGGGCCTTGCCCC 4749

Qy   1175 -------------------------GlnValThrGluGluGluAlaAlaThrAlaGlnLy 1186
Db   4750 AGAGGACAGGGACAAGCCTGCACACCAAGCAAGTGAAACAGAGCTGGCTGCGGCCATCGG 4809

Qy   1186 s---------GluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHis-- 1202
Db   4810 CTCCATCATCAATGACATTTCTGGGGAGCCAGAAAACTTCCCAGCACCTCCACCTTATCC 4869

Qy   1203 -GlyGluGlu---------------------ProGlyArgAspValLeuGluProThrGl 1215
Db   4870 TGGAGAATCCCAGACAGATCTGCAACCCCCCGCAGGTGCACAGGCGCTGCAGCCTTCTGA 4929

Qy   1215 nGlnGluLeu---ThrAlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGl 1234
Db   4930 GGAAGGAATGGAGACAGATGAGGCTGTATCTGGCATCCTGGAAACTGAGGCTGCTACAGA 4989

Qy   1234 uGlyGluValAspTrpLeuAspGlyGluLysValLysGluGluGlnGluValPheValHi 1254
Db   4990 ATCTTCTAGGCCTCCAGTCAATGCTCCTGACCCCTCA----------------------- 5026

Qy   1254 sSerGlyProAsnSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyVa 1274
Db   5027 -GCCGGCCCAACAGATACCAAGGAAGCCAGAGGAAATAGCAGTGAAACCTCACACTCAGT 5085

Qy   1274 lAlaGlyCysGlnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGl 1294
Db   5086 GCCAGAAGCCAAAGGGTCTAAAGAAGTGGAAGTCACTCTTGTTCGGAAAGACAAAGGGCG 5145

Qy   1294 uMetGluThrAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGl 1314
Db   5146 CCAGAAAACAACCCGATCACGCCGCAAGCGAAACACAAACAAGAAAGTGGTGGCT----- 5200

Qy   1314 uGlyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuTh 1334
Db   5201 ----CCTGTAGAGAGCCATGTCCCTGAATCCAACCAAGCTCAAGGT-------------- 5242

Qy   1334 rLeuAspMetProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSe 1354
Db   5243 ----GAGAGTCCTGCTGCAAATGAGGGGACAACAGTA------------CAGCACCCCGA 5286

Qy   1354 rLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrTh 1374
Db   5287 AGCCCCACAGGAAGAAAAG------------------CAGAGTGAGAAACCCCATTCCAC 5328

Qy   1374 rValThrGlnThrAla---------------------GluAlaValGluLysValIleGl 1387
Db   5329 TCCTCCTCAGTCATGTACTTCTGACCTAAGCAAGATTCCCTCCACAGAGAATTCGTCCCA 5388

Qy   1387 uThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeu----- 1405
Db   5389 AGAAATCAGTGTTGAGGAAAGGACTCCAACCAAAGCATCTGTGCCCCCAGACCTTCCCCC 5448

Qy   1406 -------LeuProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAl 1423
```

```
                    ||||||                           :::   :::
Db   5449 ACCTCCCCAGCCAGCACCGGTGGATGAGGAGCCTCAAGCCAGGTTCAGGGTGCATTCCAT 5508

Qy   1423 aGluAspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValTh 1443
              ::::::   |||:::   |||   |||         |||||||||||
Db   5509 CATTGAAAGTGACCCGGTGACCCCACCCAGCGAT---CCAAGCATCCCCATACCCACACT 5565

Qy   1443 rProAla-----------------------------------------ProGluSe 1448
            |||:::                                        |||
Db   5566 GCCTTCTGTAACTGCAGCAAAGCTCTCACCTCCTGTCGCCTCTGGGGGGATCCCACACCA 5625

Qy   1448 rThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGl 1468
            :::          :::         :::         |||::::::
Db   5626 GAGCCCCCCTACTAAGGTGACAGAGTGGATCACAAGGCAGGAGGAGCCACGGGCTCAGTC 5685

Qy   1468 uGluAspLysProAspAlaGlyProAspAlaAspGlyLysGlu-----------SerTh 1484
                   |||         ||||||         :::                 |||::
Db   5686 TACTCCATCTCCAGCTCTTCCCCCAGACACAAAGGCCTCTGATGTTGACACCAGCTCCAG 5745

Qy   1484 rAlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAs 1504
           :   :::   |||:::|||     :::|||:::   :::         |||    :::::
Db   5746 CACCCTGAGGAAGATTCTC---ATGGACCCCAAGTATGTGTCTGCCACAAGTGTCACTTC 5802

Qy   1504 nLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAl 1524
           :        |||         |||         |||                       ||
Db   5803 CACAAGTGTCACCACAGCCATTGCAGAGCCTGTC------------------AGTGCTGC 5844

Qy   1524 aProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerAr 1544
           ||||      |||                :::   |||         :::||||||::
Db   5845 CCCTTGCCTACAT------------------GAGGCCCCGCCCCCGCCAGTTGACTCTAA 5886

Qy   1544 gGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGl 1564
           ::::|||                      :::       |||       ||||||
Db   5887 AAAGCCT---------------------TTAGAAGAAAAAACAGCACCTCCAGTGACAAA 5925

Qy   1564 nProArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCy 1584
                   ::::::|||         ||||||    |||   |||         |||
Db   5926 C---AACTCTGAGATACAAGCCTCGGAGGTGCTGGTAGCT---GCTGACAAGGAAAAGGT 5979

Qy   1584 sLeuProArgLeuGlnLeuLys--------------------AlaProValSer 1595
                |||   :::      |||                        |||||||||
Db   5980 GGCTCCAGTCATTGCTCCCAAAATTACCTCTGTTATTAGCCGGATGCCTGTCAGC 6034
```

RESULT 9
US-09-917-800A-1505
; Sequence 1505, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Elashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1505
;   LENGTH: 15231
;   TYPE: DNA
;   ORGANISM: Rattus norvegicus
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020119462A1 X59601
US-09-917-800A-1505

Alignment Scores:

Pred. No.:	1.63e-20	Length:	15231
Score:	427.00	Matches:	366
Percent Similarity:	33.95%	Conservative:	258
Best Local Similarity:	19.91%	Mismatches:	745
Query Match:	5.29%	Indels:	471
DB:	10	Gaps:	70

US-09-902-432-4 (1-1596) x US-09-917-800A-1505 (1-15231)

```
Qy      3 AlaGlySerSerThrGluGlnArgSerProGluGlnProAlaGlySerAspThrProSer 22
          ||||||   :::|||      |||                  |||:::      |||
Db   4467 GCTGGCTGAGCAACAGCGGGCAGA------------------GGAGCGGGAGCGCCTGGC 4508

Qy     23 GluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspPro 42
                            |||   |||   ||||||:::   |||      |||
Db   4509 CGA------------GGTGGAGGCCGCGCTGGAGAAGCAGCGGCAGCTGGCTGAGGCCCA 4556

Qy     43 -AlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAs 62
            |||   |||         :::|||      :::         |||
Db   4557 TGCCCAGGCCAAGGCACAGGCCGAGCTGGAGGCACGAGAACTGCAGCGGCGCATGCAGGA 4616

Qy     62 nGlyValAla------GluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGl 80
              |||         |||:::      |||      |||::::::   :::      ||
Db   4617 GGAGGTGACGCGGCGCGAGGAGGCGGCGGTGGACGCACAGCAACAGAAGCGCAGCATCCA 4676

Qy     80 nGluGlu------------------------------------------GluValValAs 86
           ||||||                                          :::
Db   4677 AGAGGAGCTGCAGCATCTGCGGCAAAGCTCAGAGGCAGAGATCCAGGCCAAGGCCCAGCA 4736

Qy     86 pGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArgValGluGluMe 106
              :::      :::|||            :::   |||:::   ||||||      ::
Db   4737 GGTGGAGGCTGCAGAGCGCAGCCGCATGCGCATTGAGGAAGAGATCCGCGTAGTCCGTCT 4796

Qy    106 tAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGluGluThrSerGl 126
           :         :::|||         :::      :::   |||   |||:::      ::
Db   4797 GCAGCTAGAGACAACT---------GAGCGTCAGCGTGGAGGGGCGGAGGATGAGCTGCA 4847

Qy    126 uIleIle------------GluGlnIleProAlaSerGluAsnAsnValGluGluMetVa 142
           :   :::            |||:::      |||   :::         :::|||
Db   4848 GGCTCTGCGTGCACGGGCTGAGGAGGCAGAAGCACAGAAGCGGCAGGCTCAGGAGGAAGC 4907

Qy    142 lGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGl 162
            :::            |||      |||            :::|||
Db   4908 CGAGCGCTTGCGGAGGCAGGTGCAGGATGAGAGCCAACGCAAACGGCAGGCGGAGGCCGA 4967

Qy    162 yPheLysPheThrValLysLysAsp---------LysAsnGluLysSerAspThrValGl 179
                          ||||||   :::                   ||||||         :::||
Db   4968 GCTGGCCCTGCGTGTGAAGGCAGAAGCGGAGGCAGCGCGAGAGAAGCAGCGGGCCCTGCA 5027

Qy    179 nLeuLeuThrValLysLysAspGluGlyGluGlyAlaGlu---------------AlaSe 194
           |   |||         |||   :::   |||   ||||||                  |||
Db   5028 GGCTCTGGATGAACTGAAACTGCAGGCCGAGGAGGCCGAACGGTGGCTGTGCCAAGCCGA 5087

Qy    194 rValGlyAlaGlyAspHisGlnGluProSerValGluThrAlaValGly----------- 210
                           :::   ::::::|||||||||
Db   5088 GGCAGAGAGGGCTCGCCAAGTGCAGGTAGCCCTGGAGACAGCGCAGCGTAGTGCAGAAGT 5147

Qy    211 -GluSerAlaSerLysGluSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLe 230
            |||      ||||||               :::   |||   :::      |||   |||||
```

```
Db   5148 GGAGCTGCAGAGCAAGCGTCCGTCCTTTGCAGAGAAGACCGCACAGTTGGAGCGCACGCT 5207

Qy    230 uLysGlnGluGln---SerSerThrGluIleProLeuGlnAlaGlu---------SerAs 246
Db   5208 GCAGGAAGAGCACGTGACAGTGACACAGCTGCGGGAGGAGGCGGAACGGCGGGCACAGCA 5267

Qy    246 pGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGluProTh 266
Db   5268 GCAGGCTGAAGCCGAGCGAGCCCGTGAGGAAGCCGAGCGGGAGCTGGAGCGCTGGCAGCT 5327

Qy    266 rLysSerProGluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLy 286
Db   5328 GAAGGCCAATGAGGCGCTGCGGCTGCGGCTGCAGGCAGAGGAGGTGGCACAGCAGAAGAG 5387

Qy    286 sPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGl 306
Db   5388 CCTG-------------------------GCCCAGGCCGATGCGGAGAAGCAGAAGGA 5420

Qy    306 uAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGl 326
Db   5421 AGAGGCAGAACGGGAAGCCCGGCGGCGGGGCAAGGCAGAGGAGCAGGCCGTGCGGCAGCG 5480

Qy    326 uGluLysGluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGl 346
Db   5481 AGAGCTGGCTGAGCAGGAG---CTGGAGAAGCAGCGGCAGCTGACAGAGGGCACCGCCCA 5537

Qy    346 nAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGluAspGlnValGlyAs 366
Db   5538 GCAGCGCCTGGCTGCCGAGCAGGAGCTGATTCGCCTGCGGGCAGAGACGGAGCAAGGTGA 5597

Qy    366 pLeuGluAlaSer---SerGluGluLysCysAlaProLeu---------AlaThrGluVa 382
Db   5598 GCATCAGCGGCAGCTGCTGGAGGAAGAGCTGGCCCGGCTACAGCACGAAGCGACAGCAGC 5657

Qy    382 lPheAspGluLysMetGluAlaHisGlnGlu---------ValValAlaGluValHisVa 399
Db   5658 CACACAGAAGCGCCAGGAGCTGGAGGCTGAGCTGGCGAAGGTTCGGGCAGAGATGGAGGT 5717

Qy    399 lSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyVa 419
Db   5718 ACTGCTGGCCAGCAAGGCACGAGCCGAAGAGGAGTCTCGCTCCACCAGTGAA-------- 5769

Qy    419 lValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluVa 439
Db   5770 -----------------------------------------AAGTCCAAGCAGAGGCT 5786

Qy    439 lProGlnGluAla-----------GluProAlaGluGluLeuMetLysSerArgGluMe 455
Db   5787 GGAAGCTGAGGCAGGGCGGTTTCGAGAGCTGGCTGAGGAGGCTGCCCGCCTGCGTGCTCT 5846

Qy    455 tCysValSerGlyGlyAspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLe 475
Db   5847 GGCCGAGGAGGCAAGGCGGCACCGGGAGTTGGCC------------GAGGAGGACGCGGC 5894

Qy    475 uProLysHisProGlu------GlyIleValSerGlu---ValGluMetLeuSerSerGl 492
Db   5895 ACGCCAGCGGGCCGAGGCGGACGGAGTGCTTACGGAGAAGCTGGCTGCCATCAGTGAGGC 5954

Qy    492 nGluArgIleLys--------------------------------ValGlnGlySerPr 501
Db   5955 CACAAGGCTCAAGACGGAGGCAGAGATTGCACTCAAAGAGAAGGAGGCCGAGAACGAGCG 6014

Qy    501 oLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGl 521
Db   6015 CCTGAGGCGCCTGGCTGAAGATGAGGCCTTCCAGCGGCGCCGGCTGGAGGAGCAGGCAGC 6074

Qy    521 yLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSe 541
Db   6075 ACAGCACAAGGCAGACATAGAGGAGCGCCTGGCCCAG---------CTGCGCAAGGCATC 6125

Qy    541 rProGluSerAlaAspGluGlnLysGly------GluSerSerAlaSerSerProGluGl 559
Db   6126 CGAGAGCGAGCTGGAGCGACAGAAGGGGTTGGTGGAGGATACCCTGCGGCAGCGGCGGCA 6185

Qy    559 uProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAl 579
Db   6186 GGTGGAGGAGGAGATCATGGCTCTGAAGGCGAGCTTCGAGAAGGCCGCGGCTGGCAAGGC 6245

Qy    579 aGlu-----------------------GluGlyThrThrSerAspGlyGluLysLysAr 591
Db   6246 AGAACTGGAGCTGGAGCTTGGCCGCATCCGCAGCAATGCCGAGGACACCATGCGCAGCAA 6305

Qy    591 gGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArgValAr 611
Db   6306 GGAGCTGGCCGAGCAGGAGGCAGCGCGGCAGCGGCAGTTGGCAGCTGAGGAGGAGCAGAG 6365

Qy    611 gArgProSerGluSerAspLys--------------------GluGluGluLeuGluLy 624
Db   6366 GCGCCGGGAAGCCGAGGAGCGGGTGCAGAGGAGCCTGGCAGCGGAGGAGGAAGCCGCACG 6425

Qy    624 sValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGluVa 644
Db   6426 GCAGCGCAAGGTCGCACTGGAGGAAGTCGAGCGGCTCAAGGCCAAGGTT---GAGGAAGC 6482

Qy    644 lLysThrValGlyGluGluGlnLysProGluGluProLys------------------- 657
Db   6483 GCGGCGCCTGCGAGAGCGAGCTGAGCAGGAGTCTGCGAGGCAGCTGCAGCTGGCCCAGGA 6542

Qy    658 ----------ArgArgValAspThrSerValSerTrpGluAlaLeuIleCys-------- 671
Db   6543 GGCTGCCCAGAAACGGCTGCAGGCGGAGGAGAAGGCGCACGCCTTTGTGGTGCAGCAGCG 6602

Qy    672 --------------------------------------------------ValGlySe 674
Db   6603 AGAAGAGGAGCTGCAGCAGACTCTTCAGCAAGAGCAGAACATGCTGGAGCGGCTGCGGAG 6662

Qy    674 rSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGly---------GlyPr 691
Db   6663 CGAGGCAGAGGCAGCGCGGCGAGCTGCTGAGGAGGCGGAGGAGGCCCGGGAGCAGGCAGA 6722

Qy    691 oArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAlaGl 711
Db   6723 ACGTGAGGCAGCGCAGTCTAGGAAGCAAGTGGAAGAGGCCGAGCGGCTGAAGCAGTCG-- 6780

Qy    711 yThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGlnGlySerSerSerPr 731
Db   6781 ----------------GCAGAGGAGCAGGCTCAGGCCCAGGCCCAGGCGCAGGCGGCTGC 6824

Qy    731 oGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysAr 751
Db   6825 AGAG-------------------------------------------------------- 6828

Qy    751 gLeuValThrProArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSe 771
Db   6829 ----------------AAACTGCGCAAGGAAGCGGAGCAGGAGGCGGCGCGTCGGGCCCA 6872

Qy    771 rValGluGlnLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLy 791
Db   6873 GGCGGAGCAGGCTGCGTTGAAACAGAAGCAGGCAGCCGACGCGGAGATGGAGAAGCACAA 6932

Qy    791 sLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrVa 811
Db   6933 GAAGTTTGCAGAGCAGACGCTACGGCAGAAGGCTCAGGTAGAGCAGGAGCTGACCACGCT 6992

Qy    811 lGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAlaValValPr 831
Db   6993 GAGG-----------CTGCAGCTCGAGGAGACCGAC---CACCAGAAGAGCATCCTGGA 7037

Qy    831 oLeuSerGluTyrAsnAlaValGluArgGluLysMetGlu---AlaGlnGlyAsnThrGl 850
Db   7038 T---GAGGAGCTGCAGCGGCTAAAGGCTGAGGTAACAGAGGCAGCCCGGCAGCGTAGCCA 7094

Qy    850 uLeuProGlnLeuLeuGlyAlaValTyrValSer---GluGluLeuSerLysThrLeuVa 869
Db   7095 GGTAGAGGAGGAGCTCTTCTCTGTCCGCGTGCAGATGGAGGAGCTGGGCAAA-------- 7146
```

```
Qy     869 lHisThrValSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluAr 889
                                             :::|||            |||   ||
Db    7147 ------------------------------CTCAAGGCTCGCATTGAAGCTGAAAACCG 7175

Qy     889 gSerProSerTrpIleSerAlaSerValThrGluPro---LeuGluHisThrAlaGlyGl 908
           |:::                        |||:::      ||||||      |||   ::
Db    7176 GGCACTCATCCTTCGTGACAAGGACAACACACAGCGCTTCCTGGAGGAGGAGGCCGAGAA 7235

Qy     908 uAlaMetProProValGluGluValThrGluLysAspIleIleAlaGluGluThrProVa 928
           :               ||||||                ::: |||:::|||
Db    7236 GATGAAACAGGTGGCAGAGGAAGCTGCACGGTTGAGCGTAGCTGCCCAGGAGGCAGCAAG 7295

Qy     928 lLeuThrGln------------ThrLeuProGluGlyLysAspAlaHisAspAspMetVa 944
            |||   |||               |||   :::   :::         :::   |||::
Db    7296 GCTGCGGCAGCTAGCCGAGGAGGACCTGGCCCAGCAGCGGGCCCTGGCGGAGAAGATGCT 7355

Qy     944 lThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLe 964
           :      ::::::            ||||||                  :::|||   ||
Db    7356 GAAGGAGAAGATGCAGGCGGTGCAGGAAGCCACAAGGCTCAAGGCTGAGGCTGAGCTGCT 7415

Qy     964 uArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAl 984
           |:::   ::::::                  |||:::|||            :::   ||
Db    7416 GCAGCAGCAGAAGGAGCTG-----------GCACAGGAGCAGGCCCGGCGGCTGCAGGC 7463

Qy     984 aValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluVa 1004
           |                        |||      :::   |||      :::   |||
Db    7464 G-----------------------GACAAGGAGCAAATGGCTCAGCAGTTGGTAGAGGA 7499

Qy    1004 lGluSerGlyValLeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAl 1024
                  |||         |||   |||   ||||||||    |||      :::   ||
Db    7500 GACACAGGGTTTCAGCGGACCCTGGAGGCTGAGCGGCAGCGGCAGCTAGAAATGAGCGC 7559

Qy    1024 aValAlaAspLysValLys--------------------------------------- 1030
           |   |||:::::::|||
Db    7560 AGAGGCTGAACGCCTCAAGTTGCGCATGGCTGAGATGAGCCGGGCTCAGGCCCGTGCAGA 7619

Qy    1031 -GluGluSerGln--------------------------------------------- 1034
            |||:::::|||
Db    7620 GGAGGATGCCCAGCGCTTCCGGAAGCAGGCTGAAGAGATCGGCGAAAAGCTGCACCGCAC 7679

Qy    1035 ----------------------ValProAlaThrGlnThr------ValGlnArgThrGl 1045
                                 |||         ||||||      :::||||||
Db    7680 TGAACTCGCTACACAGGAGAAGGTGACATTGGTGCAGACTCTCGAGATCCAGCGACAGCA 7739

Qy    1045 ySer---------------LysAlaLeuGluLysValGluGluValGluGluAspSerGl 1060
            |||               :::      |||   :::   |||:::|||   :::   ||
Db    7740 GAGTGACCAAGATGCCGAGCGTCTGAGGGAGGCCATTGCTGAGCTGGAGCGTGAGAAGGA 7799

Qy    1060 u--------------------------ValLeuAlaSerGluLysGluLysAspValMe 1071
           |                              |||   ||||||:::   :::   |||
Db    7800 GAAGCTCAAGCAGGAGGCGAAGTTACTGCAGCTCAAGTCTGAGGAGATGCAGACTGTGCA 7859

Qy    1071 tProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGl 1091
               :::      :::||||||      |||         :::      ||||||
Db    7860 GCAGGAGCAGATACTGCAGGAGACACAGGCCCTGCAGAAGAGCTTTCTCTCTGAGAAGGA 7919

Qy    1091 yGlnAlaThrProGluSerLeuGluValProGluValThrAlaAspValAspHisValAl 1111
                                    :::   :::      |||   ::::::   :::
Db    7920 CAGCTTGCTGCAACGCGAACGCTTCATCGAGCAGGAGAAGGCCAAGCTGGAGCAGCTTTT 7979

Qy    1111 aThrCysGlnValIleLysLeuGlnGlnLeuMetGlu------------------GlnAl 1125
                  :::|||   |||   :::||||||   |||                  |||
Db    7980 CCAGGACGAGGTGGCAAAAGCAAAACAGCTGCAGGAGGAGCAGCAGCGGCAGCAGCAGCA 8039

Qy    1125 aValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThrProLe 1145
            :::      |||      |||   :::         |||
Db    8040 GATGGAGCAGGAAAAGCAGGAGCTGGTGGCCAGCATGGAG---------GAGGCCCGGAG 8090

Qy    1145 uAlaAspSerAspThrAlaAspGlyThr------------------------------ 1154
                    :::      :::|||
Db    8091 GCGGCAGCGTGAGGCAGAGGAGGGTGTGAGGCGCAAGCAAGAGGAACTGCAGCGTCTGGA 8150

Qy    1155 ----------GlnGlnAspGluThrIleAspSerGlnAspSerLysAlaThrAlaAlaVa 1171
                     ||||||  |||   :::   :::::::::::::::
Db    8151 GCAGCAGCGGCAGCAGCAGGAGAAACTACTGGCAGAGGAGAACCAGAGGCTGCGGGAGCG 8210

Qy    1171 lArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerTh 1191
               |||           |||      ||||||   |||     ||||||   :::||
Db    8211 GCTGCAGCGCCTGGAGGAAGAGCACCGAGCTGCGTTGGCGCACTCTGAGGAGATCGCCAC 8270

Qy    1191 r-----------------------LeuProAsnAsnValProAlaGlnGluGluHisGl 1203
           |                       |||||||||
Db    8271 CTCCCAGGCTGCTGCCACAAAAGCACTGCCCAAT------------------------- 8304

Qy    1203 yGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaVa 1223
                     |||||||||  |||:::           |||
Db    8305 ----------GGCCGCGACGCACTTGATGGCCCCTCCATGGAG---------------- 8337

Qy    1223 lProValLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGl 1243
                     |||:::  |||         ||||||           :::      ::
Db    8338 ---------GCCGAGCCCGAGTACACCTTTGAGGGATTACGTCAGAAGGTGCCAGCTCA 8387

Qy    1243 uLysValLysGlu------------GluGlnGluValPheValHisSerGlyProAsnSe 1259
           ::::::::::|||                 |||
Db    8388 GCAGCTACAGGAAGCAGGCATTCTGAGCATGGAGGAACTGCAGCGTTTGACACAGGGTCA 8447

Qy    1259 rGlnLysAlaAlaAspValThrTyrAspSerGluVal--------------------Me 1272
                     |||::::::|||        :::|||
Db    8448 CACCACGGTGGCTGAGCTCACGCAGCGGGAAGATGTGCGCCACTACCTGAAGGGCGGCAG 8507

Qy    1272 tGlyValAlaGlyCysGlnGluLysGluSerThrGluValGlnSer------------- 1287
               :::||||||        |||   :::   |||     |||
Db    8508 CAGCATCGCAGGATTGCTCCTGAAGCCCACCAATGAGAAACTGAGTGTCTACACAGCCCT 8567

Qy    1288 ------------------------------LeuSerLeuGluGlyGluMetGluTh 1297
                                         |||   |||  |||   :::      ::
Db    8568 ACAGCGGCAGCTGCTCAGCCCTGGAACAGCCCTTATCTTACTTGAGGCCCAGGCAGCCTC 8627

Qy    1297 r------------AspValGluLysGluLysArgGluThrLysProGluGlnValSerGl 1313
           :            |||     :::   :::|||   |||     |||   |||   ||
Db    8628 GGGCTTCCTGCTGGACCCTGTCCGGAACCGGCGGCTGACGGTCAATGAGGCTGTGAAGGA 8687

Qy    1313 u--------------------------------------------------------- 1313
           |
Db    8688 GGGTGTGGTGGGTCCCGAGCTGCACCACAAGCTGCTGTCAGCTGAGCGTGCCGTCACTGG 8747

Qy    1314 ----------------GluGlyGluGlnGluThrAlaAlaProGluHisGlu-------- 1325
                             |||||||||   :::     |||   ||||||
Db    8748 CTACAAGGACCCTTACACAGGAGAACAGAT-CTCTCTCTTCCAGGCCATGAAGAAGGACC 8806

Qy    1326 ----------GlyThrTyrGlyLysProValLeuThrLeuAsp---------------- 1336
                      |||    :::   |||   :::   |||
Db    8807 TCATTGTCAGGGACCATGGCATCCGCCTGCTGGAAGCCCAGATCGCCACAGGTGGCATCA 8866

Qy    1337 -------------MetProSerSerGluArgGlyLysAlaLeuGlySerLeu-------- 1349
                          |||         ||||||:::  |||   :::
Db    8867 TTGACCCTGTACACAGCCACCGTGTTCCCGTGGACGTGGCCTACCAGCGTGGCTACTTCG 8926

Qy    1350 ----GlyGlySerProSer-Leu------ProAspGlnAspLysAlaGlyCysIleGluV 1366
               |||      |||   |||     |||      |||     |||      :::
Db    8927 ATGAGGAGATGAACCGTGTGCTGGCTGACCCAAGCGATGACACCAAGGGCTTCTTTGACC 8986

Qy    1366 alGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValI 1386
                                                          |||   :::
Db    8987 CCAACACTCAC---------------------------------------GAGAACCTCA 9007

Qy    1386 leGluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuL 1406
                      ::::::          |||   ||||||  |||         |||
Db    9008 CGTACCTGCAGCTGCTGGAGCGCTGTGTGGAGGACCCCGAG---ACAGGCCTGCGCCTCC 9064

Qy    1406 euProAlaGluLysSerSerAlaThrGlyGlyHis-Trp---ThrLeuGlnHisAlaGlu 1424
           |||||            :::|||  ||||||    |||  ||||||
Db    9065 TGCCACTCACAGACAAGGCTGCCAAGGGTGGTGAGCTGGTGTACACTGACACGGAGGCCC 9124

Qy    1425 AspThrValProLeuGlyProGluSerGlnAlaGluSerIleProIleIleValThrPro 1444
              |||           |||:::            |||                 |||
```

```
Db    9125 GTGACGTCTTCGAAAAGGCCACAGTGTCTGCACCATTCGGCAAGTTCCAGGGCAAGACCG 9184

Qy    1445 AlaProGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGlu 1464
              |||                                      |||::::::|||
Db    9185 TGACCATCTGGGAGA--------------------------TCATCAACTCAGAGTACT 9217

Qy    1465 ArgSerGluGluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThr 1484
                          |||     :::|||     |||   |||:::
Db    9218 TCACAGCGGAGCAGCGACGGGACCTGCTCCGGCAGTTCCGCACG-GGCCGC------ATC 9270

Qy    1485 AlaIleGluLysValLeuLysAlaGluProGluIleLeuGluLeuGluSerLysSerAsn 1504
              :::|||||||::::::|||          ::::::|||        :::
Db    9271 ACGGTGGAGAAGATCATCAAGATTGTCATCACGGTGGTAGAGGAACACGAGCGGAAGGGC 9330

Qy    1505 LysIle-ValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAl 1524
           ::::::    |||
Db    9331 CAGCTCTGCTTTG----------------------------------------------- 9343

Qy    1524 aProGluThrHisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerAr 1544
                                    :::      ||||||   |||   |||   |||
Db    9344 ------------------------AGGGCCTCCGTGCCCTTGTGCCTGCTGCAGAGC-- 9376

Qy    1544 gGluProAsnArgCysTrpThrLysMetLysAspAlaLysMetLys 1559
                           |||||||||   ::::::   :::   ||||||
Db    9377 -------------TGCTGGACA---GTGGAGTCATCAGTCATGAAG 9406

RESULT 10
US-09-954-456-1601
; Sequence 1601, Application US/09954456
; Patent No. US20020115057A1
; GENERAL INFORMATION:
;  APPLICANT: Young, Paul
;  TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Canc
;  TITLE OF INVENTION:  Sets
;  FILE REFERENCE: 689290-76
;  CURRENT APPLICATION NUMBER: US/09/954,456
;  CURRENT FILING DATE:  2001-09-18
;  PRIOR APPLICATION NUMBER: US/60/233,617
;  PRIOR FILING DATE: 2000-09-18
;  PRIOR APPLICATION NUMBER: US/60/234,052
;  PRIOR FILING DATE: 2000-09-20
;  PRIOR APPLICATION NUMBER: US/60/234,923
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,134
;  PRIOR FILING DATE: 2000-09-25
;  PRIOR APPLICATION NUMBER: US/60/235,637
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,638
;  PRIOR FILING DATE: 2000-09-26
;  PRIOR APPLICATION NUMBER: US/60/235,711
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,720
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,840
;  PRIOR FILING DATE: 2000-09-27
;  PRIOR APPLICATION NUMBER: US/60/235,863
;  PRIOR FILING DATE: 2000-09-27
;  NUMBER OF SEQ ID NOS: 2276
;  SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1601
;   LENGTH: 14800
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-954-456-1601

Alignment Scores:
Pred. No.:                6.73e-20      Length:         14800
Score:                    418.00        Matches:        334
Percent Similarity:       35.65%        Conservative:   255
Best Local Similarity:    20.22%        Mismatches:     701
Query Match:              5.18%         Indels:         364
DB:                       10            Gaps:           61

US-09-902-432-4 (1-1596) x US-09-954-456-1601 (1-14800)

Qy       5 SerSerThrGluGlnArgSerProGlu-GlnProAlaGlySerAspThrProSerGluLe 24
           ||||||:::   ||||||||                  |||               |||
Db    4187 AGCAGCAGCGGGCAGAGGAGCGCGAGCGGCTGGCCGAGGTGGAGGCCGCGCTGGAGAAGC 4246

Qy      24 uValLeuSerGlyHisGly------ProAlaAlaGluAlaSerGlyAlaAlaGlyAspPr 42
                  :::|||   |||          |||          ||||||   |||:::
Db    4247 AGCGGCAGCTGGCCGAGGCGCACGCCCAGGCAAAGGCACAGGCGGAGCGGGAGGCGAAGG 4306

Qy      42 oAlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAs 62
            |||   |||            |||                  |||
Db    4307 AGCTGCAGCAGCGCATGCA------------GGAGGAGGTGGTGCGGCGGGAGGAGGCGG 4354

Qy      62 nGly-ValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGlnGluG 82
            |||    |||:::|||               :::|||||||||   |||:::   ::::::
Db    4355 CGGTGGACGCGCAGCAGCAGAAGCGCAGCATTCAGGAGGAGCTGCAGCAGCTGCGGCAGA 4414

Qy      82 luGluValValAspGluAspValGlyGlnArgGluSerGlu------------------ 95
                         :::             |||::::::|||
Db    4415 GCTCGGAGGCGGAGATCCAGGCCAAGGCCCGGCAGGCAGAGGCGGCTGAGCGCAGCCGGC 4474

Qy      96 --AspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluA 115
               :::   |||:::   ||||||         :::         :::|||         :::
Db    4475 TGCGCATCGAGGAGGAGATCCGCGTGGTGCGCCTGCAGTTGGAGGCCACCGAGCGCCAGC 4534

Qy     115 spIleThrLysAspGlyGlnGluGluThrSerGluIle---IleGluGlnIleProAlaS 134
                     :::|||:::   :::                  |||:::      |||
Db    4535 GTGGCGGGGCTGAGGGGGAGCTGCAGGCACTGCGTGCACGGGCGGAGGAGGCTGAGGCAC 4594

Qy     134 erGluAsnAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyP 154
             :::         :::|||      :::                  |||      |||
Db    4595 AAAAGCGACAGGCGCAGGAGGAGGCCGAGCGCTTGCGGAGGCAGGTGCAGGACGAGAGCC 4654

Qy     154 heLysLysValPheLysPheValGlyPheLysPheThrValLysLysAsp---------L 171
             :::|||            |||               ||||||   :::
Db    4655 AGCGTAAGCGGCAGGCGGAGGTGGAGCTGGCCTCGCGCGTGAAGGCCGAGACCGAGGCGG 4714

Qy     171 ysAsnGluLysSerAspThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyA 191
                 ||||||         :::|||   |||         :::   :::   |||   |
Db    4715 CGCGCGAGAAGCAGCGGGCCCTGCAGGCCCTGGAGGAGCTGCGGCTGCAGGCGGAGGAGG 4774

Qy     191 laGlu---------------AlaSerValGlyAlaGlyAspHisGlnGluProSerValG 206
           |||||               |||   |||                  :::   ::::::|
Db    4775 CGGAGCGGCGCCTGCGGCAGGCCGAGGTGGAGCGAGCGCGGCAGGTACAGGTGGCCCTGG 4834

Qy     206 luThr------------AlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnS 222
           |||||            |||      |||      ||||||   :::         :::
Db    4835 AGACGGCGCAGCGCAGTGCAGAGGCGGAGCTGCAGAGCAAACGCGCCTCCTTCGCCGAGA 4894

Qy     222 erThrGluLysGlnGluGlyThrLeuLysGlnGluGln---SerSerThrGluIleProL 241
             |||   :::   |||   :::|||::::::|||         :::      ::::::
Db    4895 AGACGGCACAGCTGGAGCGCTCCCTGCAGGAGGAACACGTGGCTGTGGCACAGCTGCGGG 4954

Qy     241 euGlnAlaGlu---------SerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyG 258
             :::||||||                  ||||||      |||   |||::::::|||   |
Db    4955 AGGAGGCTGAGCGGCGGGCACAGCAGCAGGCCGAGGCCGAGCGGGCGCGCGAGGAGGCAG 5014

Qy     258 luGluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnS 278
           ||   :::   |||:::            |||:::   |||:::              :::   :
Db    5015 AGCGGGAGCTGGAGCGCTGGCAGCTCAAGGCCAACGAGGCGCTACGGCTGCGGCTGCAGG 5074

Qy     278 erGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysL 298
           ::|||      :::      |||                              :
Db    5075 CGGAGGAGGTGGCGCAGCAGAAGAGCCTG--------------------------GCGC 5107

Qy     298 ysThrSerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluG 318
           ::         :::|||   ||||||:::               |||   :::|||   :::
Db    5108 AGGCCGAGGCTGAGAAGCAGAAGGAGGAGGCGGAGCGCGAGGCGCGGCGGCGCGGCAAGG 5167

Qy     318 lnGluAlaGluLysValAspGluGluGluLysGluLysThrGluProAlaSerGluGluG 338
             |||   :::   |||   :::   |||         :::   |||         ::::::
Db    5168 CGGAGGAGCAGGCCGTCCGGCAGCGGGAGCTGGCTGAACAAGAG---CTGGAGAAGCAGC 5224
```

```
Qy     338 lnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluL 358
Db    5225 GGCAGCTGGCGGAAGGCACCGCGCAGCAGCGCCTGGCCGCGGAGCAGGAGTTGATCCGGC 5284

Qy     358 euProLeuGluAspGlnValGlyAspLeuGluAlaSer---SerGluGluLysCysAlaP 377
Db    5285 TGCGGGCCGAGACGGAGCAGGGGGAGCAGCAGCGGCAGCTGCTGGAGGAGGAGCTGGCCC 5344

Qy     377 roLeu------------------AlaThrGluValPheAspGlu---------------- 385
Db    5345 GGCTGCAGCGTGAGGCGGCTGCAGCCACGCAGAAACGGCAGGAGCTGGAAGCCGAGCTGG 5404

Qy     386 --LysMetGluAlaHisGlnGluValVal--------------------AlaGluValH 398
Db    5405 CCAAGGTGCGGGCCGAGATGGAGGTGCTGCTGGCCAGCAAGGCGAGGGCTGAGGAGGAGT 5464

Qy     398 isValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyG 418
Db    5465 CGCGCTCCACCAGCGAGAAGTCCAAGCAGAGGCTG------------GAGGCCGAGGCCG 5512

Qy     418 lyValValValGluGlyThrGlyGlu---------------SerLeuProProGlu---- 431
Db    5513 GCCGGTTCCGCGAGCTGGCCGAGGAGGCCGCCCGCCTGCGTGCCCTGGCGGAAGAGGCCA 5572

Qy     432 -----------LysLeuAlaGluProGlnGluValProGlnGluAlaGluProAlaGluG 448
Db    5573 AGCGGCAGCGGCAGCTGGCCGAGGAAGACGCGGCGCGGCAGCGGGCCGAG---GCGGAGC 5629

Qy     448 luLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGlnLeuThrAspL 468
Db    5630 GGGTGCTTGCGGAGAAGCTGGCCGCCATCGGC---------------GAGGCCACGCGGC 5674

Qy     468 euSerProGluGluLysThrLeuProLysHisProGluGlyIleValSerGluValGluM 488
Db    5675 TCAAGACGGAGGCGGAGATCGCGCTCAAGGAGAAGGAGGCGGAGAACGAGCGCCTGCGGC 5734

Qy     488 etLeuSerSerGlnGlu---------------ArgIleLysValGlnGlySerProLeuL 503
Db    5735 GGCTGGCGGAGGACGAGGCCTTCCAGCGGCGGCGGCTGGAGGAGCAGGCCGCGCAACACA 5794

Qy     503 ysLysLeuPheSerSer------SerGlyLeuLysLysLeuSerGly---------LysL 518
Db    5795 AGGCTGACATCGAGGAGCGCCTGGCCCAGCTGCGCAAGGCATCGGACAGCGAGCTGGAGC 5854

Qy     518 ysGlnLysGlyLysArgGlyGly--------------------GlyGlyAspGluGluP 531
Db    5855 GGCAGAAGGGGCT-GGTGGAGGACACGCTGAGGCAGCGGCGGCAGGTGGAGGAAGAGATC 5913

Qy     531 roGlyGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGln------L 549
Db    5914 CTGGC------------GCTGAAGGCGAGCTTCGAGAAGGCGGCCGCTGGCAAGGCGGAG 5961

Qy     549 ysGlyGluSerSerAlaSerSerProGluGlu-ProGluGluThr---ThrCysLeuGlu 567
Db    5962 CTGGAGCTGGAGCTGGGACGCATCCGCAGCAACGCGGAGGACACGCTGCGCAGCAAGGAG 6021

Qy     568 LysGlyProLeuGluAlaProGlnAsp--------------------------------- 576
Db    6022 CAGGCCGAGCTGGAGGCCGCGAGGCAGCGGCAGCTGGCGGCGGAGGAGGAGCGGCGGCGC 6081

Qy     577 GlyGluAlaGluGluGlyThrThrSerAsp---------------------GlyGluLys 589
Db    6082 CGTGAGGCTGAGGAGCGCGTGCAGAAGAGCCTGGCGGCCGAGGAGGAGGCCGCACGGCAG 6141

Qy     590 LysArgGluGlyIleThrProTrpAlaSerPheLysLysMetValThrProLysLysArg 609
Db    6142 CGGAAGGCGGCGCTGGAGGAAGTCGAGCGGCTGAAAGCCAACGTGGAGGAGGCGCGGCGC 6201

Qy     610 ValArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThr 629
Db    6202 CTGCGGGAGCGAGCGGAGCAGGAGTCGGCGCGGCAGCTGCAGCTGGCCCAGGAGGCCGCC 6261

Qy     630 LeuSerSerThrAspSerThrValSerGluMetGlnAspGluValLysThrValGlyGlu 649
Db    6262 CAGAAGCGGCTG---------------CAGGCGGAAGAGAAGGCACACGCCTTCGCGGTG 6306

Qy     650 GluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGluAlaLeu 669
Db    6307 CAGCAGAAGGAGCAGGAGCTACAGCAGACGCTGCAGCAGGAGCAGAGC------GTGCTG 6360

Qy     670 IleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGly 689
Db    6361 GACCAGCTGCGCGGCGAGGCGGAGGCGGCCCGGCGGGCGGCTGAGGAGGCGGAGGAGGCC 6420

Qy     690 ---------GlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLys 706
Db    6421 CGGGTGCAGGCGGAGCGTGAGGCGGCGCAGGCCCGGCGGCAGGTGGAAGAGGCCGAGCGG 6480

Qy     707 AspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGlnAspGlnAlaGln 726
Db    6481 CTGAAGCAGTCG------------------GCAGAGGAGCAGGCACAGGCCCGGGCTCAG 6522

Qy     727 GlySerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrp 746
Db    6523 GCACAGGCGGCTGCAGAG------------------------------------------ 6540

Qy     747 GluSerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLys 766
Db    6541 ------------------------------AAGCTGCGCAAGGAGGCCGAGCAAGAGGCG 6570

Qy     767 AlaGluAspSerSerValGluGlnLeuSerThrGluIleGluProSerArgGluGluSer 786
Db    6571 GCGCGGCGGGCACAGGCGGAGCAGGCGGCCCTGCGGCAGAAGCAGGCAGCTGACGCGGAG 6630

Qy     787 TrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGln 806
Db    6631 ATGGAGAAGCATAAGAAATTCGCCGAGCAGACGCTGCGGCAGAAGGCGCAGGTGGAGCAG 6690

Qy     807 GluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnVal 826
Db    6691 GAGCTGACAACACTGCGG------------CTGCAGCTGGAGGAGACCGAC--------- 6729

Qy     827 ProAlaValValProLeuSerGluTyrAsnAlaValGluArgGlu-----------Lys 842
Db    6730 ------------------CACCAGAAGAACCTGCTGGACGAGGAGCTGCAGCGGCTGAAG 6771

Qy     843 MetGlu-----------AlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaVal 858
Db    6772 GCGGAGGCCACGGAGGCCGCACGCCAGCGCAGCCAGGTGGAGGAGGAGCTCTTCTCGGTG 6831

Qy     859 TyrValSer---GluGluLeuSerLysThrLeuValHisThrValSerValAlaValIle 877
Db    6832 CGCGTGCAGATGGAGGAGCTGAGCAAG------------------------------- 6858

Qy     878 AspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSer 897
Db    6859 ------CTCAAGGCACGCATCGAGGCTGAGAACCGCGCACTCATCTTGCGTGACAAGGAC 6912

Qy     898 ValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProVal---GluGluVal 916
Db    6913 AATACGCAGCGCTTCCTGCAGGAGGAGGCTGAGAAGATGAAGCAGGTGGCGGAGGAGGCC 6972

Qy     917 ThrGluLysAspIleIleAlaGluGluThrProValLeuThrGln-----------Thr 932
Db    6973 GCGCGGCTGAGTGTGGCGGCCCAAGAGGCTGCGCGACTGCGGCAGCTGGCAGAGGAGGAC 7032

Qy     933 LeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAspPheThrSer 952
Db    7033 CTGGCACAGCAGCGGGCCTTGGCAGAGAAGATGCTCAAGGAGAAGATGCAGGCGGTGCAG 7092

Qy     953 GluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAla 972
Db    7093 GAGGCCACGCGACTCAAGGCTGAGGCGGAACTGCTGCAGCAGCAGAAGGAGCTTGCGCAG 7152

Qy     973 SerGlyAlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerPro 992
```

```
Db    7153 GAGCAGGCGCGGCGGCTGCAGGAG---------------------------------- 7176

Qy     993 AspThrThrGluGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGlu 1012

Db    7177 GACAAGGAGCAGATGGCGCAGCAGCTGGCGGAGGAGACGCAGGGCTTCCAGCGGACGCTG 7236

Qy    1013 GluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLys------ 1030

Db    7237 GAGGCCGAGCGGCAGCGGCAGCTGGAGATGAGCGCTGAGGCTGAGCGCCTCAAGCTGCGT 7296

Qy    1031 ------------GluGluSerGlnValProAlaThrGlnThrValGlnArgThrGlySer 1046

Db    7297 GTGGCCGAGATGAGCCGAGCCCAGGCCCGCGCTGAGGAGGACGCCCAGCGCTTCCGGAAG 7356

Qy    1047 LysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGluLys 1066

Db    7357 CAGGCGGAGGAGATCGGTGAGAAGCTGCACCGCACGGAG------CTCGCCACCCAGGAG 7410

Qy    1067 GluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGln 1086

Db    7411 AAGGTGACCCTGGTGCAGACACTGGAGATCCAGCGACAGCAGAGTGACCATGATGCCGAG 7470

Qy    1087 GlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrAlaAsp 1106

Db    7471 ----------------------------------CGCCTGCGGGAGGCCATCGCTGAG 7494

Qy    1107 ValAspHis------ValAlaThrCysGlnValIleLysLeuGlnGlnLeu---MetGlu 1123

Db    7495 CTGGAGCGTGAGAAGGAGAAGCTCCAACAGGAGGCCAAACTGCTGCAGCTCAAGTCTGAG 7554

Qy    1124 GlnAlaValAlaProGluSerSerGluThrLeuThrAspSerGluThrAsnGlySerThr 1143

Db    7555 GAGATGCAGACGGTGCAGCAGGAGCAGCTGCTGCAGGAGACGCAGGCCCTGCAGCAAAGC 7614

Qy    1144 ProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGln 1163

Db    7615 TTCCTCTCTGAAAAGGACAGCCTG------CTACAGCGGGAGCGCTTCATCGAGCAGGAG 7668

Qy    1164 AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThr 1183

Db    7669 AAGGCCAAGCTG-----------------GAGCAGCTCTTCCAGGACGAGGTGGCCAAG 7710

Qy    1184 AlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGly 1203

Db    7711 GCACAGCAG-------------------------------CTGCGTGAGGAG------ 7731

Qy    1204 GluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThrAlaAlaAlaVal 1223

Db    7732 CAGCAGCGGCAGCAGCAGCAGATGGAGCAGGAACGGCAGCGGCTGGTGGCCAGCATGGAG 7791

Qy    1224 ProValLeuAlaLysThrGluValGlyGlnGluGly-----------------GluVal 1237

Db    7792 GAGGCGCGGCGGCGGCAGCATGAGGCCGAGGAGGGCGTGCGGCGCAAGCAGGAGGAGCTG 7851

Qy    1238 AspTrpLeuAspGlyGluLysValLysGluGluGlnGluValPheValHisSerGlyPro 1257

Db    7852 CAGCAGCTGGAGCAGCAGCGG---CGGCAGCAGGAGGAGCTGCTGGCTGAGGAGAACCAG 7908

Qy    1258 AsnSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCys 1277

Db    7909 AGGCTGCGTGAGCAGCTGCAGCTCCTGGAGGAGCAGCACCGGGCCGCGCTGGCGCACTCA 7968

Qy    1278 GlnGluLysGluSerThrGluValGlnSerLeuSer----------------------- 1289

Db    7969 GAGGAGGTCACTGCCTCGCAGGTGGCTGCCACAAAGACCCTGCCCAATGGCCGGGATGCA 8028

Qy    1290 ------------LeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGlu 1305

Db    8029 CTTGATGGCCCCGCGGCAGAGGCAGAGCCGGAGCACAGCTTCGATGGCCTGCGGCGGAAG 8088

Qy    1306 ThrLysProGluGlnValSerGluGluGlyGluGlnGluThr----------------- 1319

Db    8089 GTGTCAGCTCAGAGGCTGCAGGAGGCCGGCAT-CCTGAGTGCGGAGGAGCTGCAGCGGTT 8147

Qy    1320 ---AlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetPro 1338

Db    8148 GGCGCAGGGCCACACCACGGTGGA---------------------------------- 8171

Qy    1339 SerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAspGln 1358

Db    8172 ---------CGAGCTCGCACGGCGGGAAGACGTGCGCCACTACCTGCAGGGCCGCAGCAG 8222

Qy    1359 AspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThr 1378

Db    8223 TATCGCAGGGCTGTTGCTGAAGGCCAC-----------------------CAATGAGAA 8258

Qy    1379 AlaGluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerPro 1398

Db    8259 GCTGAGTGTTTACGCCGCCCTGCAGAGGCAGCTGCTGAGTCCCGGCACGGCCCTCATCCT 8318

Qy    1399 Glu------------------------------------------------------- 1399

Db    8319 GCTGGAGGCGCAGGCGGCCTCAGGCTTCCTGCTGGACCCTGTGCGGAACCGGCGGCTGAC 8378

Qy    1400 -----------------------CysValGlyAlaHisLeuLeuProAlaGluLysSer 1411

Db    8379 CGTCAACGAGGCTGTGAAGGAGGGTGTGGTGGGCCCCGAGCTGCACCACAAGCTGCTGTC 8438

Qy    1412 SerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyPro 1431

Db    8439 GGCCGAGCGCGCCGTCACTGGCTACAAGGACCCCTACACTGGCCAGCAGATCTCTCTCTT 8498

Qy    1432 GluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeuHis 1451

Db    8499 CCAAGCCATGCAGAAGGG-----------------CCTCATCGTCCGGGAGCACGGCAT 8540

Qy    1452 ProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAspLys 1471

Db    8541 CCG------------------------------------------------------- 8543

Qy    1472 ProAspAlaGlyProAspAlaAspGlyLys 1481

Db    8544 CCTGCTGGAGGCCCAGATCGCCACGGGCGG 8573
```

RESULT 11
US-09-815-242-4580
; Sequence 4580, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308

```
;  PRIOR FILING DATE: 2001-02-16
;  NUMBER OF SEQ ID NOS: 14110
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4580
;   LENGTH: 7104
;   TYPE: DNA
;   ORGANISM: Staphylococcus aureus
US-09-815-242-4580

Alignment Scores:
Pred. No.:              8.79e-20      Length:        7104
Score:                  411.00        Matches:       325
Percent Similarity:     35.01%        Conservative:  302
Best Local Similarity:  18.15%        Mismatches:    671
Query Match:            5.09%         Indels:        495
DB:                     10            Gaps:          72

US-09-902-432-4 (1-1596) x US-09-815-242-4580 (1-7104)

Qy     47 ProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAsnGlyValAlaGlu 66
Db   2029 CCAACAGTTAAACCTCAAGCGAAACAAGATATTATCCAAGCAGTTACAACTCGTAAACAA 2088

Qy     67 GlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGlnGluGlu---GluValVal 85
Db   2089 CAA------------ATTAAAAAGTCAAATGCATCATTACAAGATGAAAAAGATGTAGCA 2136

Qy     86 AspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAsp----------- 101
Db   2137 AATGATAAAATTGGTAAAATTGAAACAAAGGCAATTAAAGATATTGATGCAGCAACAACA 2196

Qy    102 ArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGln 121
Db   2197 AATGCACAAGTAGAAGCCATTAAAACAAAAGCGATCAATGATATTAATCAAACTGCACCT 2256

Qy    122 --------------------------GluGluThrSerGluIle------------Ile 128
Db   2257 GCTACAACAGCTAAAGCAGCAGCTCTTGAAGAATTTGACGAAGTTGTTCAAGCACAAATT 2316

Qy    129 GluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnProAlaGluSerGln 148
Db   2317 GATCAAGCACCTTTAAATCCTGATACAACAAATGAAGAAGTAGCGGAAGCTATTGAACGT 2376

Qy    149 AlaAsn------AspValGlyPheLysLysValPheLysPheValGlyPheLysPheThr 166
Db   2377 ATTAATGCAGCTAAAGTTTCTGGTGTTAAAGCAATTGAAGCGACAACGACTGCACAAGAT 2436

Qy    167 ValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr------------ 182
Db   2437 TTAGAAAGAGTTAAAAACGAAGAAATCTCAAAAATTGAAAATATTACTGACTCTACTCAA 2496

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db   2497 ACAAAAATGGATGCCTATAATGAAGTTAAACAAGCAGCAACAGCTAGAAAAACTCAAAAT 2556

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db   2557 GCTACAGTTTCAAATGCAACTAATGAAGAAGTAGCAGAAGCTGATGCAGCAGTAGAAGCA 2616

Qy    223 ThrGluLys-----------------GlnGluGlyThrLeuLysGlnGluGlnSerSer 236
Db   2617 GCTCAAAAGCAAGGTTTACATGACATCCAAGTTGTTAAATCAAAACAGGAAGTTGCTGAT 2676

Qy    237 ThrGlu-------------------------IleProLeuGlnAlaGluSerAspGln 247
Db   2677 ACAAAATCAAAAGTATTAGATAAAATCAATGCAATTCAAACACAAGCAAAAGTTAAACCT 2736

Qy    248 AlaAlaGluGluGluAlaLysAsp---------------------------------- 255
Db   2737 GCAGCTGATACGGAAGTAGAAAACGCATATAATACACGTAAACAAGAAATTCAAAATAGC 2796

Qy    256 ---------GluGlyGluGluLysGln---------GluLysGluProThrLysSerPro 269
Db   2797 AATGCTTCAACTACAGAAGAAAAACAAGCTGCATATACAGAATTAGATACTAAAAAGCAA 2856

Qy    270 GluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThr 289
Db   2857 GAAGCAAGAACAAATCTTGATGCTGCAAATACAAACAGTGATGTAACAACAGCTAAAGAC 2916

Qy    290 HisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAsp---Asp 308
Db   2917 AATGGTATTGCTGCAATTAATCAAGTACAAGCGGCAACAACTAAGAAATCGGATGCTAAA 2976

Qy    309 LeuGluThrAlaGluLysArgLysGluGlnGluAla------GluLysValAspGluGlu 326
Db   2977 GCGGAAATCGCTCAAAAAGCAAGTGAACGTAAAACTGCAATTGAAGCAATGAATGATTCG 3036

Qy    327 GluLysGluLysThrGluProAlaSerGluGluGlnGluProAla-------------- 341
Db   3037 ACTACTGAAGAACAACAAGCTGCAAAAGATAAAGTTGATCAAGCAGTAGTTACTGCAAAC 3096

Qy    342 GluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGlu 361
Db   3097 GCTGATATAGATAATGCTGCAGCAAATACTGATGTAGATAATGCAAAAACTACTAATGAA 3156

Qy    362 AspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGlu 381
Db   3157 GCTACAATCGCAGCCATTACACCTGATGCAAATGTTAAA------CCAACAGCGAAACAA 3210

Qy    382 ValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSerThr 401
Db   3211 GCAATTGCTGATAAAGTACAAGCGCAAGAAACAGCAATT-------------------- 3249

Qy    402 ValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValVal 421
Db   3250 ------------------------------------GATGCTAATAACGGTGCAACAACA 3273

Qy    422 GluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGln 441
Db   3274 GAA----------------------GAAAAAGCAGCTGCGAAACAACAAGTTCAAACT 3309

Qy    442 GluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAsp 461
Db   3310 GAAAAAACAACAGCTGATACAGCAATT----------------------GATGGTGCA 3345

Qy    462 HisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGly 481
Db   3346 CATACAAAT-----------------GCAGAAGTTGAAGCGGCTAAAAATGCAGAA--- 3384

Qy    482 IleValSerGluValGluMetLeu-SerSerGlnGluArgIleLysValGlnGlySerPr 501
Db   3385 ---ATTGCTAAAATTGAAGCAATTCAGCCAGCAACAACAACTAAAGATAATGCGAAAC-- 3439

Qy    501 oLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGl 521
Db   3440 ----AAGCAATTGCTACGA-----------AAGCGAATGAACGTAAAACAGCAATCGCT 3483

Qy    521 yLysArgGlyGlyGlyGlyAspGluGlu-----------ProGlyGluTyrGlnHisIl 537
Db   3484 CAAACGCAAGACATT-ACTGCTGAAGAAATTGCAGCGGCAAATGCGAATGTAGATAATGC 3542

Qy    537 eHisThrGlu-------------------------SerProGluSerAlaAspGluGl 548
Db   3543 TGTGACACAAGCAAATAACAACATTGAAGCTGCTAATAGTCAAAATGATGTAGACCAAGC 3602

Qy    548 nLys------GlyGluSerSerAlaSer------SerProGluGluProGluGluThrTh 564
Db   3603 GAAAACAACTGGTGAAGCTAGTATTGATCAAGTAACACCAACAGTTAATAAAAAAGCAAC 3662

Qy    564 rCysLeu------GluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGl 582
Db   3663 AGCAGTTACAGACGCGAAAAATAATATTACAGCTGCAACTGATGATAATGGTGTAGATAC 3722

Qy    582 yThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLy 602
Db   3723 AGCGAAAGATGCAGGTAAAAATTCGATTCAAAGTACACAACCAGCAACAGCGGTTAAATC 3782
```

```
Qy      602 s---------------------------MetValThrProLysLysArgValArgArgPr 613
                                        :::   |||      :::   :::
Db     3783 AAATGCTAAAAATGATGTTGATCAAGCTGTGACAACTCAAAATCAAGCAATTGATAATAC 3842

Qy      613 oSerGluSerAspLysGluGlu-------------------------------------- 620
             :::   :::      ||||||
Db     3843 AACTGGTGCTACAACTGAAGAAAAAAATGCAGCAAAAGATTTAGTTTTAAAAGCTAAAGA 3902

Qy      621 ----------------------------------------GluLeuGluLysValLysSe 627
                                                    ::::::   :::::::||| 
Db     3903 AAAAGCGTATCAAGATATCTTAAATGCACAAACAACTAACGATGTTACGCAAATTAAAGA 3962

Qy      627 rAlaThrLeuSer---------------SerThrAspSerThrValSerGluMet---Gl 641
                   ::::::               :::   |||:::|||:::   ::::::   ::
Db     3963 TCAAGCAGTTGCTGATGTTCAAGGTATTACTGCAGATACAACAATTAAAGATGTTGCGAA 4022

Qy      641 nAspGluValLysThrValGlyGluGluGlnLys----------------------------- 652
            :||||||:::   |||         |||||||||
Db     4023 AGATGAATTAGCAACAAAAGCAAGGGAACAAAAAGCACTTATTGCACAAACTGCAGATGC 4082

Qy      653 -------------ProGluGluProLysArgArgValAspThrSerValSerTrp----- 666
                            |||:::      ::::::||||||      ::::::
Db     4083 GACTACTGAAGAAAAAGAACAAGCAAATCAACAAGTAGATGCAGAATTAACGCAAGGTAA 4142

Qy      667 -------------------GluAlaLeuIleCysValGlySerSerLysLysArgAlaAr 680
                               :::::::::      |||   ::::::|||      |||
Db     4143 TCAAAATATTGAAAATGCACAGTCAATCGATGATGTAAACACTGCAAAAGATAATGCAAT 4202

Qy      680 gLysAla---------------SerSerSerAspAspGluGlyGlyProArgThr----- 693
             :::|||               :::|||:::|||   :::         |||
Db     4203 TCAAGCAATTGACCCAATTCAAGCATCAACAGATGTTAAAACGAATGCAAGAGCGGAATT 4262

Qy      694 ----------------------------------LeuGlyGlyAspSerHisArgAlaGl 702
                                              |||      :::            ||
Db     4263 GCTAACTGAAATGCAAAATAAAATAACTGAAATACTTAATAATAATGAGACTACTAATGA 4322

Qy      702 uGluAlaSerLysAsp---------------------LysGluAlaGlyThrAspAlaVa 715
            ||||         |||                        |||   |||   :::   ::
Db     4323 AGAAAAAGGTAACGATATTGGACCAGTTAGAGCTGCATATGAAGAAGGTTTAAATAATAT 4382

Qy      715 lProAlaSerThr---------------------------------------GlnGluGl 722
            :   |||:::|||                                       |||:::
Db     4383 TAATGCAGCAACTACTACAGGTGATGTAACTACTGCTAAAGATACAGCAGTACAAAAAGT 4442

Qy      722 nAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGlyGluGl 742
                |||         :::         :::|||||||||      :::      :::
Db     4443 TCAACAACTTCATGCAAATCCTGTTAAGAAGCCAGCAGGTAAAACTGCTTTAGATCAAGC 4502

Qy      742 yValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLysSerLy 762
                :::      :::      :::      :::||||||            :::
Db     4503 TGCAGCTGAT---AAGAAAACACAAATAGAACAAACACCAAATGCGTCACAACAAGAAAT 4559

Qy      762 sLeuGluGluLysAlaGlu---AspSerSerValGluGlnLeuSerThrGluIleGluPr 781
                :::   |||   |||   |||:::   :::   |||      |||   |||:::
Db     4560 TAACGATGCAAAACAAGAAGTTGATACTGAATTAAATCAAGCGAAAACAAATATTGACCA 4619

Qy      781 oSerArgGluGluSerTrpVal------------------------------------Se 789
             |||      :::   :::|||                                    ::
Db     4620 ATCTTCAACAGATGAATATGTTGATAATGCAGTTAAAGAAGGAAAAGCTAAAATTAATGC 4679

Qy      789 rIleLysLysPheIleProGlyArgArg-------------------------------- 798
            ::::|||   |||         ::::::
Db     4680 AGTTAAAACATTTAGTGAGTACAAAAAAGATGCTTTAGCTAAGATTGAAGCTGCATATAA 4739

Qy      798 ------------------------------------------------------------ 798

Db     4740 TGCTAAAGTAACCGAAGCGGATAACTCTAACGCATCGACTTCAAGTGAAATTGCAGAAGC 4799

Qy      799 ----------------------LysLysArgAlaAspGlyLysGlnGluGlnAlaThrVa 811
                                  |||:::   ||||||            |||||||||
Db     4800 GAAACAAAAACTTGCTGAATTGAAACAAACGGCGGATCAAAATGTCAATCAAGCTACTTC 4859

Qy      811 lGluAspSerGlyProValGluIleAsnGluAspAspProAsnVal-------------- 826
             :::|||         |||:::|||:::   |||      |||:::
Db     4860 TAAAGATGACATTGAAGTTCAAATTCATAATGACTTAGATAATATTAACGATTACACAAT 4919

Qy      827 -------------------ProAlaValValProLeuSerGluTyrAsnAlaValGluAr 840
                                  |||         |||   |||          :::::
Db     4920 TCCAACAGGTAAAAAAGAATCAGCTACAACAGATTTATATGCCTATGCAGATCAGAAGAA 4979

Qy      840 gGluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVa 860
            :      :::   |||      |||            :::         |||:::
Db     4980 AAATAATATTTCAGCTGACACTAATGCAACACAAGATGAAAAGCAACAAGCAATTAAGCA 5039

Qy      860 lSerGluGluLeuSerLysThrLeuValHisThrValSer-------------------- 873
                ::::::      :::|||   :::   :::::::::
Db     5040 AGTTGACCAAAATGTTCAAACTGCATTAGAAAGCATTAATAATGGTGTGGATAATGGTGA 5099

Qy      874 -------------------------ValAlaValIleAspGlyThrArgAlaValThrSe 885
                                        |||   ||||||      :::         ::
Db     5100 CGTTGATGATGCATTAACACAAGGTAAAGCAGCAATTGATGCTATTCAAGTAGATGCTAC 5159

Qy      885 rValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGlu----- 903
            :|||:::   ::::::         |||   |||         |||      |||
Db     5160 TGTTAAACCTAAAGCGAACCAAGTGATTGATGCTAAAGCAGAGGAAACGAAAGAATCTAT 5219

Qy      904 -------------------------HisThrAlaGlyGluAlaMetProProValGluGl 915
                                                 ||||||:::      ::::::::
Db     5220 TGATCAAAGTGACCAGTTAACTGCTGAAGAAAAAACTGAAGCATTAGCAATGATTAAACA 5279

Qy      915 uValThrGlu------LysAspIleIleAlaGluGluThrProVal------LeuThrGl 931
            ::::|||:::      |||      |||   :::   |||            :::   ::
Db     5280 AATTACAGATCAAGCTAAACAAGGTATTACTGATGCAACAACAACTGCTGAAGTTGAAAA 5339

Qy      931 nThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAspPheTh 951
            :         :::|||   :::|||   |||::::::         ::::::|||   ||
Db     5340 AGCGAAAGCTCAAGGACTTGAGGCATTTGATAACATT---------CAAATTGACTCAAC 5390

Qy      951 rSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGl 971
            |   :::                  |||      ||||||         |||
Db     5391 AGAAAAACAAAAAGCTATCGAAGAATTAGAAACTGCACTAGACCAGATTGAAGCAGGTGT 5450

Qy      971 uAlaSerGlyAlaGluGluThrThrAsp-------------------------------- 980
                      |||:::   ||||||:::
Db     5451 AAATGTCGACGCTGATGCTACAACTGAAGAAAAAGAAGCGTTTACGAATGCTTTAGAAGA 5510

Qy      981 ----MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaTh 999
                :::      |||         :::::|||                |||   |||||
Db     5511 CATTTTATCAAAAGCAACTGAAGATATTTCTGATCAAACTACAAATGCAGAAATCGCTAC 5570

Qy      999 rProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGl 1019
            |            |||::::::      |||:::      :::   :::|||
Db     5571 T------------GTCAAAAATAGTGCGCTTGAACAACTTAAAGCACAACGTATTAATCC 5618

Qy     1019 nAlaIle------------LeuGlnAlaValAlaAspLysValLysGluGluSerGlnVa 1035
                :::            |||:::|||::::::      |||   ::::::   :::::
Db     5619 AGTAGTTAAGAAAAATGCTTTGGAAGCAATTAGAGAAGTGGTTAACAAGCAAATAGAAAT 5678

Qy     1035 l---------------------ProAlaThrGlnThrValGlnArgThrGly-------- 1045
            :                        :::   :::   :::   ||||||
Db     5679 AATTAAAAATGCAGATGCAGATGCATCGGCGAAAGAAATTGCACGTACAGATTTAGGTAG 5738

Qy     1046 -------SerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAl 1063
                      :::      :::|||:::::::::   :::   ::::::||||||
Db     5739 ATATTTTGACAGATTTGCTGATAAATTAGATAAAACACAAACAAATACAGAAGTT----- 5793

Qy     1063 aSerGluLysGluLysAspValMetProLys------------GlyProValGlnGluAl 1079
             :::|||   :::         :::|||                  |||      :::
Db     5794 -GCTGAATTACAAAATGTAACGATACCTGCAATTGAAGCGATTGTGCCTCAGAATGATCC 5852

Qy     1079 aGlyAlaGluHisLeuAlaGlnGlySerGlu------------------------ThrGl 1091
                |||               ||||||:::
Db     5853 AAATGCTAATGATACTAATAGTGGTTCAGATAATAATGATGCAACAGCAAATTCAAATGC 5912

Qy     1091 yGlnAlaThrProGluSerLeuGluValProGluValThr---------------AlaAs 1106
```

```
                  |||||||||||:::           |||   ||||||              |||::
Db    5913 TAATGCAACACCAGAAAATACTGGACAACCTAATGTAACTGAGTCAACAGATAATGCCAA 5972

Qy    1106 pValAspHisValAlaThrCys-GlnValIleLysLeu--------------------- 1118
            :   |||       :::|||    |||:::|||::::::
Db    5973 TGCAGATACGTCTTCAACAACGACAAATAATCAGAATGATGCGGCAACAGGTGAAACAAC 6032

Qy    1119 --------------------GlnGlnLeuMetGluGlnAlaValAlaProGluSerSerG 1132
                               |||||||||||   |||    :::   |||:::::::::
Db    6033 AGCAACATCAGCAAACAGTTCAGCAACTGATGATGCAAATGATAAA-CCGCAAGCTAATA 6091

Qy    1132 luThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaA 1152
              :::   ::::::|||:::   ::::::|||        |||:::|||
Db    6092 ACAAT---TCTAGTGCAGATACATCAACAAACAGTCCGACAATGGATAATGATGTAACTA 6148

Qy    1152 spGlyThrGlnGlnAspGluThrIleAsp---SerGlnAspSerLysAlaThrAlaAlaV 1171
                 :::   :::   |||   :::   :::   |||      ||||||   |||
Db    6149 GCAAACCAGAAGTAGAATCAACTAATAACGGCACAACTGATAAACCTGCAACAGAAGCA- 6207

Qy    1171 alArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerT 1191
                             |||      |||:::||||||                 :::
Db    6208 --------GATAATGCAACGCCAGCAGAAAGCGCAACAAATAACAATAGTACAACTACAG 6259

Qy    1191 hrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValL 1211
                       |||   |||   :::
Db    6260 CAACAAACGAGAATGCACCAACAGAA---------------------------------- 6285

Qy    1211 euGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluV 1231
                                   :::   |||      |||      |||   |||
Db    6286 ----------------------TCAACAGCAACGGCTCCAACAACTGCGAGCACA---- 6318

Qy    1231 alGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu---GluGlnG 1250
              |||   |||         |||      |||         |||   :::   :::|||
Db    6319 --GGTGCAGAATCATCTGCTGATAGCAAAGATAATGCATCTGTAAATGATTCAAAACAAA 6376

Qy    1250 luValPheValHisSerGly-----------------------------------ProA 1258
                    |||::::::                                    |||
Db    6377 ATGCTGAAGTAAATAATAGTGCTGAATCTCAATCAACTAATGGCAAGGTTGCACAACCAA 6436

Qy    1258 snSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysG 1278
              |||:::      |||            |||
Db    6437 AATCTGAAAATAAAGCTAAGGCAGAAAAAGATGGTAGAGATTCAACAAATCAAAGTATGG 6496

Qy    1278 lnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrA 1298
              |||      :::      :::   |||      :::   |||   ::::::   ::::
Db    6497 TTGAATCAACAACTGAAACATTGCCTTCAGCAGACATAACAGAACCAAAAGTATCTTCTA 6556

Qy    1298 spValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu---------G 1315
            ::         |||:::|||   |||:::            |||      |||
Db    6557 ATACATCAAAAGATAAAGAAGAAAGTACGACAAGTCAAACAGATGCTGAGCAACACAATT 6616

Qy    1315 lyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGly---------------- 1329
              :::         |||:::   |||   :::   :::   |||
Db    6617 CTGATACAAACGTTGCATCTAACGAAGCAGATAAATCTGAAGGCAACGTTGATACTGATG 6676

Qy    1330 --------LysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuG 1347
                   ||||||      :::         ||||||      :::   ||||||
Db    6677 TATCGAATAAACCA------TCAACATCTAAACCGTCAGAAGCAAAAGATAAAGCG---- 6726

Qy    1347 lySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValG 1367
                                  |||      |||      ||||||
Db    6727 -----------------ACTTCGACTGAGGATAGTCAAAAAGCTGACATGGCAACAGCAG 6769

Qy    1367 lnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIleG 1387
                 :::            ::::::      ||||||:::         |||      :
Db    6770 ATACAAAAGATAATCAAGCATCAATTGGTGCAACTGCAGATGTAAACAATAAAGCTACGC 6829

Qy    1387 luThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuP 1407
            ::                                         ||||||:::      |
Db    6830 AAAAC-----------------------------------GACGGTGCCAATGCATCTC 6853

Qy    1407 roAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrV 1427
            |||||                                                ||||
Db    6854 CAGCT---------------------------------------------------ACAG 6862

Qy    1427 alProLeuGlyProGluSer---GlnAlaGluSerIleProIleIleValThrProAlaP 1446
              ||      |||      |||         :::   :::   :::      |||
Db    6863 TTTCAAAAGGAAGCAATAGTGCTAATCAAGATATGCTTAATGTAACTAAGACTAAAGAAA 6922

Qy    1446 roGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgS 1466
              ::::::                     ||||||:::::::::         :::   :::
Db    6923 ACAAAGCAAATGCAAAATCAGCCCAACAAGGAAAAGTTAAT---------AAGCCTAAAC 6973

Qy    1466 erGluGluGluAspLysProAspAlaGly 1475
              :::   :::         ||||||   |||
Db    6974 AACAAGCTAAAACTTTACCAGATACTGGT 7002
```

RESULT 12
US-09-815-242-8291
; Sequence 8291, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8291
; LENGTH: 7107
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7107)
US-09-815-242-8291

Alignment Scores:

Pred. No.:	8.8e-20	Length:	7107
Score:	411.00	Matches:	325
Percent Similarity:	35.01%	Conservative:	302
Best Local Similarity:	18.15%	Mismatches:	671
Query Match:	5.09%	Indels:	495
DB:	10	Gaps:	72

US-09-902-432-4 (1-1596) x US-09-815-242-8291 (1-7107)

```
Qy      47 ProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSerSerValAsnGlyValAlaGlu 66
              |||      |||         |||         :::   :::|||            :::
Db    2029 CCAACAGTTAAACCTCAAGCGAAACAAGATATTATCCAAGCAGTTACAACTCGTAAACAA 2088

Qy      67 GlnGlyAspValHisValGlnGluGluAsnGlnGluGlyGlnGluGlu---GluValVal 85
           |||         :::::::::   |||         |||:::|||   :::|||
```

```
Db   2089 CAA-----------ATTAAAAAGTCAAATGCATCATTACAAGATGAAAAAGATGTAGCA 2136

Qy     86 AspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAsp----------- 101
Db   2137 AATGATAAAATTGGTAAAATTGAAACAAAGGCAATTAAAGATATTGATGCAGCAACAACA 2196

Qy    102 ArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGln 121
Db   2197 AATGCACAAGTAGAAGCCATTAAAACAAAAGCGATCAATGATATTAATCAAACTGCACCT 2256

Qy    122 --------------------------GluGluThrSerGluIle------------Ile 128
Db   2257 GCTACAACAGCTAAAGCAGCAGCTCTTGAAGAATTTGACGAAGTTGTTCAAGCACAAATT 2316

Qy    129 GluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnProAlaGluSerGln 148
Db   2317 GATCAAGCACCTTTAAATCCTGATACAACAAATGAAGAAGTAGCGGAAGCTATTGAACGT 2376

Qy    149 AlaAsn------AspValGlyPheLysLysValPheLysPheValGlyPheLysPheThr 166
Db   2377 ATTAATGCAGCTAAAGTTTCTGGTGTTAAAGCAATTGAAGCGACAACGACTGCACAAGAT 2436

Qy    167 ValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr----------- 182
Db   2437 TTAGAAAGAGTTAAAAACGAAGAAATCTCAAAAATTGAAAATATTACTGACTCTACTCAA 2496

Qy    183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
Db   2497 ACAAAAATGGATGCCTATAATGAAGTTAAACAAGCAGCAACAGCTAGAAAAACTCAAAAT 2556

Qy    203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
Db   2557 GCTACAGTTTCAAATGCAACTAATGAAGAAGTAGCAGAAGCTGATGCAGCAGTAGAAGCA 2616

Qy    223 ThrGluLys-----------------GlnGluGlyThrLeuLysGlnGluGlnSerSer 236
Db   2617 GCTCAAAAGCAAGGTTTACATGACATCCAAGTTGTTAAATCAAAACAGGAAGTTGCTGAT 2676

Qy    237 ThrGlu-------------------------IleProLeuGlnAlaGluSerAspGln 247
Db   2677 ACAAAATCAAAAGTATTAGATAAAATCAATGCAATTCAAACACAAGCAAAAGTTAAACCT 2736

Qy    248 AlaAlaGluGluGluAlaLysAsp----------------------------------- 255
Db   2737 GCAGCTGATACGGAAGTAGAAAACGCATATAATACACGTAAACAAGAAATTCAAAATAGC 2796

Qy    256 ---------GluGlyGluGluLysGln---------GluLysGluProThrLysSerPro 269
Db   2797 AATGCTTCAACTACAGAAGAAAAACAAGCTGCATATACAGAATTAGATACTAAAAAGCAA 2856

Qy    270 GluSerProSerSerProValAsnSerGluThrThrSerSerPheLysLysPhePheThr 289
Db   2857 GAAGCAAGAACAAATCTTGATGCTGCAAATACAAACAGTGATGTAACAACAGCTAAAGAC 2916

Qy    290 HisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAsp---Asp 308
Db   2917 AATGGTATTGCTGCAATTAATCAAGTACAAGCGGCAACAACTAAGAAATCGGATGCTAAA 2976

Qy    309 LeuGluThrAlaGluLysArgLysGluGlnGluAla------GluLysValAspGluGlu 326
Db   2977 GCGGAAATCGCTCAAAAAGCAAGTGAACGTAAAACTGCAATTGAAGCAATGAATGATTCG 3036

Qy    327 GluLysGluLysThrGluProAlaSerGluGluGlnGluProAla-------------- 341
Db   3037 ACTACTGAAGAACAACAAGCTGCAAAAGATAAAGTTGATCAAGCAGTAGTTACTGCAAAC 3096

Qy    342 GluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGlu 361
Db   3097 GCTGATATAGATAATGCTGCAGCAAATACTGATGTAGATAATGCAAAAACTACTAATGAA 3156

Qy    362 AspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGlu 381
Db   3157 GCTACAATCGCAGCCATTACACCTGATGCAAATGTTAAA------CCAACAGCGAAACAA 3210

Qy    382 ValPheAspGluLysMetGluAlaHisGlnGluValValAlaGluValHisValSerThr 401
Db   3211 GCAATTGCTGATAAAGTACAAGCGCAAGAAACAGCAATT-------------------- 3249

Qy    402 ValGluLysThrGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValVal 421
Db   3250 ----------------------------------GATGCTAATAACGGTGCAACAACA 3273

Qy    422 GluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGln 441
Db   3274 GAA----------------------GAAAAAGCAGCTGCGAAACAACAAGTTCAAACT 3309

Qy    442 GluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAsp 461
Db   3310 GAAAAAACAACAGCTGATACAGCAATT----------------------GATGGTGCA 3345

Qy    462 HisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGly 481
Db   3346 CATACAAAT-----------------GCAGAAGTTGAAGCGGCTAAAAATGCAGAA--- 3384

Qy    482 IleValSerGluValGluMetLeu-SerSerGlnGluArgIleLysValGlnGlySerPr 501
Db   3385 ---ATTGCTAAAATTGAAGCAATTCAGCCAGCAACAACAACTAAAGATAATGCGAAAC-- 3439

Qy    501 oLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGl 521
Db   3440 ----AAGCAATTGCTACGA-----------AAGCGAATGAACGTAAAACAGCAATCGCT 3483

Qy    521 yLysArgGlyGlyGlyGlyAspGluGlu-----------ProGlyGluTyrGlnHisIl 537
Db   3484 CAAACGCAAGACATT-ACTGCTGAAGAAATTGCAGCGGCAAATGCGAATGTAGATAATGC 3542

Qy    537 eHisThrGlu--------------------------SerProGluSerAlaAspGluGl 548
Db   3543 TGTGACACAAGCAAATAACAACATTGAAGCTGCTAATAGTCAAAATGATGTAGACCAAGC 3602

Qy    548 nLys------GlyGluSerSerAlaSer------SerProGluGluProGluGluThrTh 564
Db   3603 GAAAACAACTGGTGAAGCTAGTATTGATCAAGTAACACCAACAGTTAATAAAAAAGCAAC 3662

Qy    564 rCysLeu------GluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGl 582
Db   3663 AGCAGTTACAGACGCGAAAAATAATATTACAGCTGCAACTGATGATAATGGTGTAGATAC 3722

Qy    582 yThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLy 602
Db   3723 AGCGAAAGATGCAGGTAAAAATTCGATTCAAAGTACACAACCAGCAACAGCGGTTAAATC 3782

Qy    602 s-------------------------MetValThrProLysLysArgValArgArgPr 613
Db   3783 AAATGCTAAAAATGATGTTGATCAAGCTGTGACAACTCAAAATCAAGCAATTGATAATAC 3842

Qy    613 oSerGluSerAspLysGluGlu--------------------------------------- 620
Db   3843 AACTGGTGCTACAACTGAAGAAAAAAATGCAGCAAAAGATTTAGTTTTAAAAGCTAAAGA 3902

Qy    621 ---------------------------------------GluLeuGluLysValLysSe 627
Db   3903 AAAAGCGTATCAAGATATCTTAAATGCACAAACAACTAACGATGTTACGCAAATTAAAGA 3962

Qy    627 rAlaThrLeuSer---------------SerThrAspSerThrValSerGluMet---Gl 641
Db   3963 TCAAGCAGTTGCTGATGTTCAAGGTATTACTGCAGATACAACAATTAAAGATGTTGCGAA 4022

Qy    641 nAspGluValLysThrValGlyGluGluGlnLys------------------------- 652
Db   4023 AGATGAATTAGCAACAAAAGCAAGGGAACAAAAAGCACTTATTGCACAAACTGCAGATGC 4082

Qy    653 ------------ProGluGluProLysArgArgValAspThrSerValSerTrp----- 666
Db   4083 GACTACTGAAGAAAAAGAACAAGCAAATCAACAAGTAGATGCAGAATTAACGCAAGGTAA 4142
```

```
Qy    667 ------------------GluAlaLeuIleCysValGlySerSerLysLysArgAlaAr 680
Db   4143 TCAAAATATTGAAAATGCACAGTCAATCGATGATGTAAACACTGCAAAAGATAATGCAAT 4202
Qy    680 gLysAla---------------SerSerSerAspAspGluGlyGlyProArgThr---- 693
Db   4203 TCAAGCAATTGACCCAATTCAAGCATCAACAGATGTTAAAACGAATGCAAGAGCGGAATT 4262
Qy    694 ---------------------------------LeuGlyGlyAspSerHisArgAlaGl 702
Db   4263 GCTAACTGAAATGCAAAATAAAATAACTGAAATACTTAATAATAATGAGACTACTAATGA 4322
Qy    702 uGluAlaSerLysAsp--------------------LysGluAlaGlyThrAspAlaVa 715
Db   4323 AGAAAAAGGTAACGATATTGGACCAGTTAGAGCTGCATATGAAGAAGGTTTAAATAATAT 4382
Qy    715 lProAlaSerThr---------------------------------------GlnGluGl 722
Db   4383 TAATGCAGCAACTACTACAGGTGATGTAACTACTGCTAAAGATACAGCAGTACAAAAAGT 4442
Qy    722 nAspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGlyGluGl 742
Db   4443 TCAACAACTTCATGCAAATCCTGTTAAGAAGCCAGCAGGTAAAACTGCTTTAGATCAAGC 4502
Qy    742 yValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLysSerLy 762
Db   4503 TGCAGCTGAT---AAGAAAACACAAATAGAACAAACACCAAATGCGTCACAACAAGAAAT 4559
Qy    762 sLeuGluGluLysAlaGlu---AspSerSerValGluGlnLeuSerThrGluIleGluPr 781
Db   4560 TAACGATGCAAAACAAGAAGTTGATACTGAATTAAATCAAGCGAAAACAAATATTGACCA 4619
Qy    781 oSerArgGluGluSerTrpVal-----------------------------------Se 789
Db   4620 ATCTTCAACAGATGAATATGTTGATAATGCAGTTAAAGAAGGAAAAGCTAAAATTAATGC 4679
Qy    789 rIleLysLysPheIleProGlyArgArg------------------------------ 798
Db   4680 AGTTAAAACATTTAGTGAGTACAAAAAAGATGCTTTAGCTAAGATTGAAGCTGCATATAA 4739
Qy    798 ------------------------------------------------------------ 798
Db   4740 TGCTAAAGTAACCGAAGCGGATAACTCTAACGCATCGACTTCAAGTGAAATTGCAGAAGC 4799
Qy    799 ---------------------LysLysArgAlaAspGlyLysGlnGluGlnAlaThrVa 811
Db   4800 GAAACAAAAACTTGCTGAATTGAAACAAACGGCGGATCAAAATGTCAATCAAGCTACTTC 4859
Qy    811 lGluAspSerGlyProValGluIleAsnGluAspAspProAsnVal------------- 826
Db   4860 TAAAGATGACATTGAAGTTCAAATTCATAATGACTTAGATAATATTAACGATTACACAAT 4919
Qy    827 ------------------ProAlaValValProLeuSerGluTyrAsnAlaValGluAr 840
Db   4920 TCCAACAGGTAAAAAAGAATCAGCTACAACAGATTTATATGCCTATGCAGATCAGAAGAA 4979
Qy    840 gGluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVa 860
Db   4980 AAATAATATTTCAGCTGACACTAATGCAACACAAGATGAAAAGCAACAAGCAATTAAGCA 5039
Qy    860 lSerGluGluLeuSerLysThrLeuValHisThrValSer------------------- 873
Db   5040 AGTTGACCAAAATGTTCAAACTGCATTAGAAAGCATTAATAATGGTGTGGATAATGGTGA 5099
Qy    874 ------------------------ValAlaValIleAspGlyThrArgAlaValThrSe 885
Db   5100 CGTTGATGATGCATTAACACAAGGTAAAGCAGCAATTGATGCTATTCAAGTAGATGCTAC 5159
Qy    885 rValGluGluArgSerProSerTrpIleSerAlaSerValThrGluProLeuGlu----- 903
Db   5160 TGTTAAACCTAAAGCGAACCAAGTGATTGATGCTAAAGCAGAGGAAACGAAAGAATCTAT 5219
Qy    904 ------------------------HisThrAlaGlyGluAlaMetProProValGluGl 915

Db   5220 TGATCAAAGTGACCAGTTAACTGCTGAAGAAAAAACTGAAGCATTAGCAATGATTAAACA 5279
Qy    915 uValThrGlu------LysAspIleIleAlaGluGluThrProVal------LeuThrGl 931
Db   5280 AATTACAGATCAAGCTAAACAAGGTATTACTGATGCAACAACAACTGCTGAAGTTGAAAA 5339
Qy    931 nThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAspPheTh 951
Db   5340 AGCGAAAGCTCAAGGACTTGAGGCATTTGATAACATT---------CAAATTGACTCAAC 5390
Qy    951 rSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGl 971
Db   5391 AGAAAAACAAAAAGCTATCGAAGAATTAGAAACTGCACTAGACCAGATTGAAGCAGGTGT 5450
Qy    971 uAlaSerGlyAlaGluGluThrThrAsp------------------------------ 980
Db   5451 AAATGTCGACGCTGATGCTACAACTGAAGAAAAAGAAGCGTTTACGAATGCTTTAGAAGA 5510
Qy    981 ----MetValSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaTh 999
Db   5511 CATTTTATCAAAAGCAACTGAAGATATTTCTGATCAAACTACAAATGCAGAAATCGCTAC 5570
Qy    999 rProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGl 1019
Db   5571 T-----------GTCAAAAATAGTGCGCTTGAACAACTTAAAGCACAACGTATTAATCC 5618
Qy   1019 nAlaIle-----------LeuGlnAlaValAlaAspLysValLysGluGluSerGlnVa 1035
Db   5619 AGTAGTTAAGAAAAATGCTTTGGAAGCAATTAGAGAAGTGGTTAACAAGCAAATAGAAAT 5678
Qy   1035 l--------------------ProAlaThrGlnThrValGlnArgThrGly-------- 1045
Db   5679 AATTAAAAATGCAGATGCAGATGCATCGGCGAAAGAAATTGCACGTACAGATTTAGGTAG 5738
Qy   1046 -------SerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAl 1063
Db   5739 ATATTTTGACAGATTTGCTGATAAATTAGATAAAACACAAACAAATACAGAAGTT----- 5793
Qy   1063 aSerGluLysGluLysAspValMetProLys-----------GlyProValGlnGluAl 1079
Db   5794 -GCTGAATTACAAAATGTAACGATACCTGCAATTGAAGCGATTGTGCCTCAGAATGATCC 5852
Qy   1079 aGlyAlaGluHisLeuAlaGlnGlySerGlu----------------------ThrGl 1091
Db   5853 AAATGCTAATGATACTAATAGTGGTTCAGATAATAATGATGCAACAGCAAATTCAAATGC 5912
Qy   1091 yGlnAlaThrProGluSerLeuGluValProGluValThr--------------AlaAs 1106
Db   5913 TAATGCAACACCAGAAAATACTGGACAACCTAATGTAACTGAGTCAACAGATAATGCCAA 5972
Qy   1106 pValAspHisValAlaThrCys-GlnValIleLysLeu--------------------- 1118
Db   5973 TGCAGATACGTCTTCAACAACGACAAATAATCAGAATGATGCGGCAACAGGTGAAACAAC 6032
Qy   1119 -------------------GlnGlnLeuMetGluGlnAlaValAlaProGluSerSerG 1132
Db   6033 AGCAACATCAGCAAACAGTTCAGCAACTGATGATGCAAATGATAAA-CCGCAAGCTAATA 6091
Qy   1132 luThrLeuThrAspSerGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaA 1152
Db   6092 ACAAT---TCTAGTGCAGATACATCAACAAACAGTCCGACAATGGATAATGATGTAACTA 6148
Qy   1152 spGlyThrGlnGlnAspGluThrIleAsp---SerGlnAspSerLysAlaThrAlaAlaV 1171
Db   6149 GCAAACCAGAAGTAGAATCAACTAATAACGGCACAACTGATAAACCTGCAACAGAAGCA- 6207
Qy   1171 alArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerT 1191
Db   6208 --------GATAATGCAACGCCAGCAGAAAGCGCAACAAATAACAATAGTACAACTACAG 6259
Qy   1191 hrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValL 1211
```

```
Db   6260 CAACAAACGAGAATGCACCAACAGAA---------------------------------- 6285

Qy   1211 euGluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluV 1231
                                 :::    |||      |||      |||   |||
Db   6286 -----------------------TCAACAGCAACGGCTCCAACAACTGCGAGCACA---- 6318

Qy   1231 alGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu---GluGlnG 1250
             |||   |||         |||         |||          :::    :::|||
Db   6319 --GGTGCAGAATCATCTGCTGATAGCAAAGATAATGCATCTGTAAATGATTCAAAACAAA 6376

Qy   1250 luValPheValHisSerGly------------------------------------ProA 1258
                  |||:::::                                        |||
Db   6377 ATGCTGAAGTAAATAATAGTGCTGAATCTCAATCAACTAATGGCAAGGTTGCACAACCAA 6436

Qy   1258 snSerGlnLysAlaAlaAspValThrTyrAspSerGluValMetGlyValAlaGlyCysG 1278
             |||:::      |||            |||
Db   6437 AATCTGAAAATAAAGCTAAGGCAGAAAAAGATGGTAGAGATTCAACAAATCAAAGTATGG 6496

Qy   1278 lnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluThrA 1298
             |||      :::      :::   |||      :::    |||   ::::::  ::::
Db   6497 TTGAATCAACAACTGAAACATTGCCTTCAGCAGACATAACAGAACCAAAAGTATCTTCTA 6556

Qy   1298 spValGluLysGluLysArgGluThrLysProGluGlnValSerGluGlu---------G 1315
          ::      |||:::|||   |||:::            |||         |||
Db   6557 ATACATCAAAAGATAAAGAAGAAAGTACGACAAGTCAAACAGATGCTGAGCAACACAATT 6616

Qy   1315 lyGluGlnGluThrAlaAlaProGluHisGluGlyThrTyrGly---------------- 1329
            :::         |||:::      :::      :::    |||
Db   6617 CTGATACAAACGTTGCATCTAACGAAGCAGATAAATCTGAAGGCAACGTTGATACTGATG 6676

Qy   1330 --------LysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeuG 1347
                  ||||||      :::         ||||||      :::   ||||||
Db   6677 TATCGAATAAACCA------TCAACATCTAAACCGTCAGAAGCAAAAGATAAAGCG---- 6726

Qy   1347 lySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValG 1367
                              |||      |||      ||||||
Db   6727 -----------------ACTTCGACTGAGGATAGTCAAAAAGCTGACATGGCAACAGCAG 6769

Qy   1367 lnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIleG 1387
               :::            ::::::      ||||||:::         |||      :
Db   6770 ATACAAAAGATAATCAAGCATCAATTGGTGCAACTGCAGATGTAAACAATAAAGCTACGC 6829

Qy   1387 luThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuP 1407
          ::                                          ||||||:::      |
Db   6830 AAAAC-----------------------------------GACGGTGCCAATGCATCTC 6853

Qy   1407 roAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrV 1427
          |||||                                                   ||||
Db   6854 CAGCT---------------------------------------------------ACAG 6862

Qy   1427 alProLeuGlyProGluSer---GlnAlaGluSerIleProIleIleValThrProAlaP 1446
          ||      |||      |||          :::    :::    :::      |||
Db   6863 TTTCAAAAGGAAGCAATAGTGCTAATCAAGATATGCTTAATGTAACTAAGACTAAAGAAA 6922

Qy   1446 roGluSerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgS 1466
            ::::::                  ||||||:::::::::         :::   :::
Db   6923 ACAAAGCAAATGCAAAATCAGCCCAACAAGGAAAAGTTAAT---------AAGCCTAAAC 6973

Qy   1466 erGluGluGluAspLysProAspAlaGly 1475
            :::   :::      ||||||   |||
Db   6974 AACAAGCTAAAACTTTACCAGATACTGGT 7002
```

RESULT 13
US-09-960-253-145
; Sequence 145, Application US/09960253
; Patent No. US20020123619A1
; GENERAL INFORMATION:
; APPLICANT: Benson, Darin R.
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Lodes, Michael J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.556
; CURRENT APPLICATION NUMBER: US/09/960,253
; CURRENT FILING DATE: 2001-09-20
; NUMBER OF SEQ ID NOS: 187
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 145
; LENGTH: 10300
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-960-253-145

Alignment Scores:			
Pred. No.:	2.21e-19	Length:	10300
Score:	408.00	Matches:	315
Percent Similarity:	38.11%	Conservative:	305
Best Local Similarity:	19.36%	Mismatches:	670
Query Match:	5.05%	Indels:	337
DB:	10	Gaps:	60

US-09-902-432-4 (1-1596) x US-09-960-253-145 (1-10300)

```
Qy     26 LeuSerGly------------HisGlyProAlaAlaGluAlaSerGlyAlaAla------ 39
          ||||||||             ||||||          |||   ||||||   |||
Db   2376 CTCTCAGGTGAAGGAACTTAGCATGGTAACAGAATTGAGGGCTCAGGTAAAGCAACTGGA 2435

Qy     40 GlyAspPro-AlaAspAlaAspProAlaThrLysLeuProGlnLysAsnGlyGlnLeuSe 59
             :::|||  |||:::|||:::              :::|||        ::::::     |||     ::
Db   2436 AATGAACCTTGCAGAAGCAGAAAGGCAAAGAAGACTTGATTATGAAAGC---CAAACTGC 2492

Qy     59 rSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsnGlnGluGl 79
          :      |||   :::     ||||||            :::|||                       :::
Db   2493 CCATGACAACCTGCTCACTGAACAG------ATCCATAGTCTCAGCATAGAAGCCAAATC 2546

Qy     79 yGlnGluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGl 99
           ::::::    ::::::                ::::|||       |||    |||     :::||||||:::
Db   2547 TAAAGATGTGAAAATT---------GAAGTTTTACAGAATGAACTGGATGATGTGCAG-- 2595

Qy     99 uLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAs 119
                          :::    :::         ||||||      :::         :::
Db   2596 ------------CTTCAGTTTTCTGAGCAGAGTACCCTGATAAGAAGCCTGCAAAGCCA 2642

Qy    119 pGlyGlnGluGluThrSerGluIleIle---------GluGlnIleProAlaSerGluAs 136
              |||     :::      ||||||::::::          |||::::::                ::
Db   2643 GCTGCAAAATAAGGAAAGTGAAGTGCTTGAGGGGGCAGAACGTGTAAGGCATATCTCAAG 2702

Qy    136 nAsnValGluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLy 156
          :   |||||||||:::      |||                  :::         ::::::
Db   2703 TAAAGTGGAAGAACTGTCCCAGGCTCTTTCACAGAAGGAACTTGAAATA----------- 2751

Qy    156 sValPheLysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAs 176
                                            |||      |||
Db   2752 ------------------------------ACAAAAATGGAT------------------ 2763

Qy    176 pThrValGlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGl 196
                 |||||||||              ||||||         :::      |||        :::      ::::::
Db   2764 -------CAGCTCTTACTAGAGAAAAAGAGAGATGTGGAAACCCTCCAACAAACCATCGA 2816

Qy    196 yAlaGlyAspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGl 216
                 |||   |||            :::        ::::::     |||             :::
Db   2817 GGAGAAGGATCAACAAGTGACAGAAATCAGCTTTAGTATGACTGAGAAAATGGTTCAGCT 2876

Qy    216 uSerGluLeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSe 236
           :::|||   |||   |||         :::        |||||||||:::
Db   2877 TAATGAAGAGAAGTTTTCTCTTGGGGTTGAAATTAAGACTCTTAAAGAA----------- 2925

Qy    236 rThrGluIleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGl 256
              ::::::      |||   :::      :::::::::|||      :::|||:::         ::::::::
Db   2926 ----CAGCTAAATTTATTATCCAGAGCTGAGGAAGCAAAAAAAGAGCAGGTGGAAGAAGA 2981

Qy    256 uGlyGluGluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProVa 276
          :   |||                  :::                    ||||||:::       ::
Db   2982 TAATGAAGTTTCTTCTGGCCTTAAACAAAATTATGATGAGATGAGCCCAGCAGGACAAAT 3041

Qy    276 lAsnSerGluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpAr 296
```

301

```
                  ::::   |||                                    |||   :::          ::
Db     3042 AAGTAAGGAAGAA--------------------CTTCAGCATGAATTTGACCTTCTGAA 3080

Qy      296 gLysLysThrSerPheLysLysSerLys-----------------------GluAspAs 308
            :|||:::         :::|||  |||                        ::::::
Db     3081 GAAAGAAAATGAGCAGAGAAAGAGAAAGCTCCAGGCAGCTCTTATTAACAGAAAGGAGCT 3140

Qy      308 pLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLy 328
             |||:::       :::    :::|||:::   |||         |||||     :::||
Db     3141 TCTGCAAAGAGTCAGTAGATTGGAAGAAGAATTAGCCAACTTGAAAGATGAATCTAAGAA 3200

Qy      328 sGluLysThrGluProAlaSerGluGluGlnGluProAlaGluAspThrAspGlnAlaAr 348
            ||||             :::|||      |||     |||||   :::
Db     3201 AGAAATCCCACTCAGTGAGACTGAGAGGGGAGAAGTGGAAGAAGATAAAGAAAACAAAGA 3260

Qy      348 gLeuSer---------------AlaAspTyrGluLysValGluLeuProLeuGluAspGl 363
              |||                 :::     ::::::::|||:::    |||:::
Db     3261 ATACTCAGAAAAATGTGTGACTTCTAAGTGCCAAGAAATAGAAATTTATTTAAAACAGAC 3320

Qy      363 nValGlyAspLeuGluAla--------------------SerSerGluGluLysCysAl 376
             :::   :::   |||                        ||||||||
Db     3321 AATATCTGAGAAAGAAGTGGAACTACAGCATATAAGGAAGGATTTGGAAGAAAAG----- 3375

Qy      376 aProLeuAlaThrGluValPheAspGluLysMetGluAlaHisGlnGluValValAlaGl 396
                 ||||||   |||                 |||      |||   :::|||   ::
Db     3376 ----CTGGCAGCTGAA------------------GAGCAATTCCAGGCTCTGGTCAAACA 3413

Qy      396 uValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGluAlaGl 416
            ::::::   :::      :::||||||   :::
Db     3414 GATGAATCAGACCTTGCAAGATAAAACAAACCAA------------------------- 3447

Qy      416 uGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluPr 436
                                         |||       :::::::|||
Db     3448 ----------------------------ATAGATTTGCTCCAAGCAGAAATCAGTGAAAA 3479

Qy      436 oGlnGluValProGlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCy 456
              |||   :::   |||:::                     :::   :::
Db     3480 CCAAGCAATTATCCAGAAG--------------------TTAATCACAAGTAACACGGA 3518

Qy      456 sValSerGlyGlyAspHisThrGlnLeuThrAsp-----------LeuSerProGluGl 472
                |||    ||||||          |||                :::||||||
Db     3519 TGCAAGTGATGGGGACTCCGTAGCACTTGTAAAGGAAACAGTGGTGATAAGTCCACCTTG 3578

Qy      472 uLysThrLeuProLysHisProGluGlyIleValSerGluValGluMetLeuSerSerGl 492
                        :::|||     :::       :::   |||       :::          ::
Db     3579 TACAGGTAGTAGTGAACACTGGAAACCAGAACTAGAAGAAAAGATACTGGCCCTTGAAAA 3638

Qy      492 nGluArgIleLysValGlnGlySerProLeuLysLysLeuPheSerSerSerGlyLeuLy 512
            :|||:::    ::::::|||          |||    |||         |||          :::
Db     3639 AGAAAAGGAGCAACTTCAAAAGAAGCTACAGGAAGCCTTAACCTCCCGCAAGGCAATTCT 3698

Qy      512 sLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyGlyAspGluGluProGl 532
             |||    :::    :::|||::::::         |||            ||||||
Db     3699 TAAAAAGGCACAGGAGAAAGAAAGACATCTCAGG---------------GAGGAGCTAAA 3743

Qy      532 yGluTyrGlnHisIleHisThrGluSerProGluSerAlaAspGluGlnLysGlyGluSe 552
             :::    :::       :::             |||      ||||||||     |||::
Db     3744 GCAACAGAAAGATGACTATAATCGCTTGCAAGAACAGTTTGATGAGCAAAGCAAGGAAAA 3803

Qy      552 r---------------------------------------------------SerAl 554
            :
Db     3804 TGAGAATATTGGAGACCAGCTAAGGCAACTCCAGATTCAAGTAAGGGAATCCATAGACGG 3863

Qy      554 aSerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaPr 574
                |||           ::::::   :::|||                ||||||   ||
Db     3864 AAAACTCCCAAGCACAGACCAGCAGGAATCGTGTTCTTCCACTCCAGGTTTAGAAGAACC 3923

Qy      574 oGlnAspGlyGluAlaGluGluGlyThrThr---------------------------- 584
            |                   |||:::       |||
Db     3924 TTTATTCAAAGCCACAGAACAGCATCACACTCAACCTGTTTAGAGTCCAACTTGTGCCC 3983

Qy      585 ---------------SerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSe 599
                           |||:::                 :::|||    |||      |||
Db     3984 AGACTGGCCTTCTCATTCTGAAGATGCGAGTGCTCTGCAGGGCGGAACTTCTGTTGCCCA 4043

Qy      599 rPheLysLysMetValThrProLysLysArgValArgArgProSerGluSerAspLysGl 619
                |||      :::                                          |||||
Db     4044 GATTAAGGCCCAGCTG--------------------------------------AAGGA 4064

Qy      619 uGluGluLeuGluLysValLys---SerAlaThrLeuSerSerThrAspSerThrValSe 638
             |   |||    ||||||||:::                :::||||||||    |||    :::::
Db     4065 AATAGAGGCTGAGAAAGTAGAGTTAGAATTGAAAGTTAGTTCTACAACAAGTGAGCTTAC 4124

Qy      638 rGluMetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysAr 658
            ::::        :::||||||        :::     |||:::        :::      :::
Db     4125 TAAAAAATCAGAAGAGGTATTTCAGTTACAAGAGCAGATAAATAAACAGGGTTTAGAAAT 4184

Qy      658 gArgValAspThrSerValSerTrpGluAlaLeuIleCysValGlySerSerLysLysAr 678
                  :::||||||      ||||||     :::            |||     :::::::
Db     4185 CGAGAGTCTAAAGACAGTATCCCATGAAGCTGAAGTCCATGCCGAAAGCCTGCAGCAGAA 4244

Qy      678 gAlaArgLysAlaSerSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSe 698
            :                              |||              :::   |||
Db     4245 ATTG----------------------GAAAGCAGCCAACTACAAATTGCTGGCCTAGA 4280

Qy      698 rHisArgAlaGluGluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSe 718
              |||       |||         |||       |||                          ||
Db     4281 ACATCTAAGAGAATTGCAACCTAAACTGGATGAACTG---------CAAAAACTCATAAG 4331

Qy      718 rThrGlnGluGlnAspGlnAlaGlnGlySer---------SerSerProGluProAlaGl 735
             |   :::|||:::|||     :::           |||          |||      |||    |||
Db     4332 CAAAAAGGAAGAAGACGTTAGCTACCTTTCTGGACAACTTAGTGAGAAAGAAGCAGCTCT 4391

Qy      735 ySerProSerGluGlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrPr 755
             :::       :::   |||   :::           :::        |||    |||    |||
Db     4392 CACTAAAATACAGACAGAGATAATAGAACAAGAAGATTTAATTAAGGCTCTGCATACA-- 4449

Qy      755 oArgLysLysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGlnLe 775
                          :::   ::::::::   |||     |||        ::::::||||||
Db     4450 ------------CAGCTAGAAATGCAAGCCAAAGAGCATGATGAGAGGATAAAGCAGCT 4496

Qy      775 uSerThrGluIle--------------------GluProSerArgGluGluSerTrpVa 788
             |       |||:::                               |||          ||||||||
Db     4497 ACAGGTGGAACTTTGTGAAATGAAGCAAAAACCAGAAGAGATTGGAGAAGAAAGT----- 4551

Qy      788 lSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGl 808
                                        |||    |||::::::        |||    :::
Db     4552 ------------------------AGAGCAAAGCAACAAATACAAAGGAAACTGCAAGC 4586

Qy      808 nAlaThrValGluAspSerGlyProValGluIleAsnGluAspAspProAsnValProAl 828
              |||    :::                   |||       :::::::::    ::::::
Db     4587 TGCCCTTATTTCCCGAAAA---------GAAGCACTAAAAGAAAACAAAAGTCTCCAAGA 4637

Qy      828 aValValProLeuSerGluTyrAsnAlaValGluArg--------------------- 840
                 :::    |||:::        :::||||||
Db     4638 GGAATTGTCTTTGGCCAGA---GGTACCATTGAACGTCTCACCAAGTCTCTGGCAGATGT 4694

Qy      841 -GluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLeuLeuGlyAlaValTyrVa 860
              |||        :::        |||    |||          :::||||||     :::     ::
Db     4695 GGAAAGCCAAGTTTCTGCTCAAAATAAAGAAAAAGATACGGTCTTAGGAAGGTTAGCTCT 4754

Qy      860 lSerGluGluLeuSerLysThrLeuValHisThrValSerValAlaValIleAspGlyTh 880
            :    :::|||              |||:::       :::          :::::::::::
Db     4755 TCTTCAAGAAGAAAGAGACAAACTCATTACAGAAATGGACAGGTCTTTATTGGAAAATCA 4814

Qy      880 rArgAlaValThrSerValGluGluArgSerProSerTrpIleSerAlaSerValThrGl 900
                      :::|||     |||         |||                     :::|||||
Db     4815 GAGTCTCAGCAGCTCCTGTGAA------AGTCTAAAACTAGCTCTAGAGGGTCTTACTGA 4868

Qy      900 uProLeuGluHisThrAlaGlyGluAlaMetProProValGluGluValThrGluLysAs 920
             |        |||         |||                     |||          :::
Db     4869 AGACAAGGAAAAAGTTAGTGAAGGAAATT---------------GAATCTTTGAAATCTTC 4913

Qy      920 pIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHi 940
              |||||||||   |||                             |||    ::::::    ||
Db     4914 TAAGATTGCAGAAAGTACT-----------------------GAGTGGCAAGAGAAACA 4949
```

```
Qy    940 sAspAsp---------------------MetValThrSerGluValAspPheThrSerGl 953
Db   4950 CAAGGAGCTACAAAAAGAGTATGAAATTCTTCTGCAGTCCTATGAGAATGTTAGTAATGA 5009

Qy    953 uAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGlu-------------- 968
Db   5010 AGCAGAAAGGATTCAGCATGTGGTGGAAGCTGTGAGGCAAGAGAAACAAGAACTGTATGG 5069

Qy    969 ----------ValThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaVa 985
Db   5070 CAAGTTAAGAAGCACAGAGGCAAACAAGAAGGAGACAGAAAAGCAGTTGCAGGAAGCTGA 5129

Qy    985 lSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGlnGluValGl 1005
Db   5130 GCAAGAAATGGAGGAAATGAAAGAAAAGATGAGAAAGTTT------GCTAAATCTAAACA 5183

Qy   1005 uSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaVa 1025
Db   5184 GCAGAAAATCCTAGAGCTGGAAGAAGAGAATGACCGGCTTAGGGCAGAGGTGCACCCTGC 5243

Qy   1025 lAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrGl 1045
Db   5244 AGGAGATACAGCTAAAGAG----------------TGTATGGAAACACTTCTTCTTCCAA 5288

Qy   1045 ySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeuAlaSerGl 1065
Db   5289 TGCCAGCATGAAGGAAGAACTTGAAAGGGTCAAAATGGAGTATGAAACCCTTTCTAAGAA 5348

Qy   1065 u---------------LysGluLysAspValMetProLysGlyProValGlnGluAlaGl 1080
Db   5349 GTTTCAGTCTTTAATGTCTGAGAAAGACTCTCTAAGTGAA---GAGGTTCAAGAT----- 5400

Qy   1080 yAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluVa 1100
Db   5401 -TTAAAGCATCAGATAGAAGATAATGTATCTAAACAAGCTAACCTAGAGGCCACCGAGAA 5459

Qy   1100 lProGluValThrAlaAspValAspHisValAlaThrCysGlnVal---------IleLy 1117
Db   5460 ACATGATAACCAAACGAATGTCACTGAAGAGGGAACACAGTCTATACCAGGTGAGACTGA 5519

Qy   1117 sLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSerGluThrLeuThrAspSe 1137
Db   5520 AGAGCAAGACTCTCTGAGTATGAGCACAAGACCTACATGTTCAGAATCGGTTCCATCAGC 5579

Qy   1137 rGluThrAsnGlySerThrProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAs 1157
Db   5580 GAAGAGT------GCCAACCCTGCTGTAAGTAAGGATTTCAGCTCACATGATGAAATTAA 5633

Qy   1157 p---------GluThrIleAspSerGlnAspSerLysAlaThrAlaAlaValArgGlnSe 1174
Db   5634 TAACTACCTACAGCAGATTGATCAGCTCAAAGAAAGAATTGCTGGATTAGAGGAGGAGAA 5693

Qy   1174 rGlnValThrGluGluGluAlaAlaThrAlaGlnLysGluGluProSerThrLeuProAs 1194
Db   5694 GCAGAAAAACAAGGAATTTAGCCAGACTTTAGAAAATGAGAAA---AATACCTTACTGAG 5750

Qy   1194 nAsnValProAlaGlnGluGluHisGlyGluGluProGlyArgAspValLeuGluProTh 1214
Db   5751 TCAGATATCAACAAAGGAT------GGTGAA-------------------CTAAAATGCT 5786

Qy   1214 rGlnGlnGluLeuThr---------------------------AlaAlaAlaValProVa 1225
Db   5787 TCAGGAGGAAGTAACCAAAATGAACCTGTTAAATCAGCAAATCCAAGAAGAACTCTCCAG 5846

Qy   1225 lLeuAlaLysThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysVa 1245
Db   5847 AGTTACCAAACTAAAGGAGACAGCAGAAGAAGAGAAAGATGATTTGGAAGAGAGGCTTAT 5906

Qy   1245 lLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspVa 1265
Db   5907 GAATCAATTAGCAGAACTT---------AATGGAAGCATTGGGAATTACTGTCAGGATGT 5957

Qy   1265 lThr------------------------TyrAspSerGluValMetGlyValAlaGlyCy 1277
Db   5958 TACAGATGCCCAAATAAAAAATGAGCTATTGGAATCTGAAATGAAGAACCTTAAAAAGTG 6017

Qy   1277 sGlnGluLysGluSerThrGluValGlnSerLeuSerLeuGluGluGlyGluMetGluTh 1297
Db   6018 TGTGAGTGAATTGGAAGAAGAAAAGCAGCAGTTAGTCAAGGAAAAAACTAAGGTGGAATC 6077

Qy   1297 rAspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGl 1317
Db   6078 AGAAATACGAAAGGAATATTTGGAGAAAATACAAGGTGCTCAGAAAGAACCCGGAAATAA 6137

Qy   1317 nGluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMe 1337
Db   6138 A---------------------------AGCCATGCAAAGGAACTTCAGGAACTGTTAAA 6170

Qy   1337 tProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAs 1357
Db   6171 AGAAAAACAACAAGAAGTAAAGCAGCTA------------CAGAAGGACTGCATCAGGTA 6218

Qy   1357 pGlnAspLysAlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGl 1377
Db   6219 TCAAGAGAAAATTAGTGCTCTGGAGAGAACTGTTAAAGCTCTAGAATTTGTTCAAACTGA 6278

Qy   1377 nThrAlaGluAlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSe 1397
Db   6279 ATCTCAAAAAGATTTGGAAATAACCAAAGAAAAT------CTGGCTCAAGCAGTTGAACA 6332

Qy   1397 rProGluCysValGlyAlaHis------------------LeuLeuProAlaGluLysSe 1411
Db   6333 CCGCAAAAAGGCACAAGCAGAATTAGCTAGCTTCAAAGTCCTGCTAGATGACACTCAAAG 6392

Qy   1411 rSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGlyPr 1431
Db   6393 TGAAGCAGCA------------------AGGGTCCTAGCAGACAATCTCAAGTTGAAAAA 6434

Qy   1431 oGluSerGlnAla------GluSerIleProIleIleValThrProAlaProGluSerTh 1449
Db   6435 GGAACTTCAGTCAAATAAAGAATCAGTTAAAAGCCAGATGAAACAAAAGGATGAA----- 6489

Qy   1449 rLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGl 1469
Db   6490 ----------GATCTTGAGCGAAGACTGGAACAGGCAGAAGAGAAGCACCTGAAAGAGAA 6539

Qy   1469 uAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysVa 1489
Db   6540 GAAGAATATGCAAGAGAAACTGGATGCTTTGCGCAGAGAAAAAGTCCACTTGGAAGAGAC 6599

Qy   1489 lLeu------------------------LysAlaGluProGluIleLeuGluLeuGluSe 1501
Db   6600 AATTGGAGAGATTCAGGTTACTTTGAACAAGAAAGACAAGGAAGTTCAGCAACTTCAGGA 6659

Qy   1501 rLysSerAsnLysIleVal 1507
Db   6660 AAACTTGGACAGTACTGTG 6678
```

RESULT 14
US-09-880-107-3423
; Sequence 3423, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02

```
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3423
;   LENGTH: 4840
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 U76366
US-09-880-107-3423

Alignment Scores:
Pred. No.:              4.21e-19      Length:        4840
Score:                  398.50        Matches:       382
Percent Similarity:     34.23%        Conservative:  242
Best Local Similarity:  20.95%        Mismatches:    700
Query Match:            4.94%         Indels:        505
DB:                     10            Gaps:          78

US-09-902-432-4 (1-1596) x US-09-880-107-3423 (1-4840)

Qy      10 ArgSerProGluGlnProAlaGlySerAspThrProSerGluLeuValLeuSerGlyHis 29
           |||:::      :::||||||||                               |||:::
Db      42 CGTGCAGGTAGCCGGCCGGCCGGGGGT-----------------------CGCGGGTAT 77

Qy      30 GlyProAlaAlaGluAlaSerGlyAlaAlaGlyAspProAlaAspAlaAspProAlaThr 49
           |||         ||||||:::||||||              |||        |||:::
Db      78 GGCCGAGGCCAGGAAGCGGCGGGAGCTACTTCC------CCTGATCTACCACCATCTGCT 131

Qy      50 -LysLeuProGlnLysAsnGlyGlnLeuSerSerValAsnGlyValAlaGluGlnGlyAs 69
            :::                |||   :::   :::         |||   ||||||
Db     132 GCGGGCT------------GGCTATGTGCGTGCGGCGCGGGAAGTGAAGGAGCAGAGCGG 179

Qy      69 p---------------------------------------------ValHisValGlnGl 74
                                                         |||   |||::
Db     180 CCAGAAGTGTTTCCTGGCTCAGCCCGTAACCCTTCTGGACATCTATACACTGGCAACA 239

Qy      74 uGluAsnGlnGluGlyGlnGluGluGluValValAspGluAspValGlyGlnArgGluSe 94
           :   ::::::   |||::::::   :::       :::||||||        :::
Db     240 AACCTCAGAGCTTGGTCGGAAGCGGAAG---GCAGAGGAAGATGCGGCACTGCAA----- 291

Qy      94 rGluAspValArgGluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGl 114
                    :::|||   ||||||   :::       :::   |||   :::   ||
Db     292 ---------GCTAAGAAAACCCGTGTGTCAGACCCCATCAGCACCTCGGAGAGCTCGGA 341

Qy     114 uAspIleThrLysAspGlyGlnGluGluThrSerGluIleIleGluGlnIleProAlaSe 134
           |:::   ::::::   :::   ||||||::::::         ::::::    :::::
Db     342 AGAG---GAGGAAGAAGCAGAAGCCGAAACCGCCAAAGCCACCCCAAGACTAGCATCTAC 398

Qy     134 rGluAsnAsnVal-----------------GluGluMetValGlnProAlaGluSerGl 148
           :   ::::::|||                  |||   :::   |||::::::::
Db     399 CAACTCCTCAGTCCTGGGGGCGGACTTGCCATCAAGCATGAAAGAAAAAGCCAAGGCAGA 458

Qy     148 n--------------------AlaAsnAspValGlyPheLysLysValPheLysPheVa 161
           :                    |||   :::                  |||   ||
Db     459 GACAGAGAAAGCTGGCAAGACTGGGAATTCCATGCCACACCCTGCCACTGGAAGACGGT 518

Qy     161 lGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLe 181
           |                 :::   |||         :::            :::
Db     519 GGCCAACCTTCTTTCTGGGAAGTCTCCCAGGAAGTCAGCAGAGCCCTCAGCAAATACTAC 578

Qy     181 uThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGl 201
            |||   :::   |||   ||||||:::         :::   ||||||
Db     579 GTTGGTCTCAGAAACTGAGGAGGAGGGCAGCGTCCCGGCCTTTGGAGCTGCT------GC 632

Qy     201 nGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGl 221
            :::|||   :::   :::|||   |||::::::   |||            :::
Db     633 CAAGCCTGGGATGGTGTCAGCG---GGCCAGGCCGACAGCTCCAGCGAGGACACCTCCAG 689

Qy     221 nSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLe 241
            |||:::::::::   :::      :::|||      :::       ::::::
Db     690 CTCCAGTGATGAGACAGACGTGGAGGTAAAGGCCTCTGAAAAAATTCTCCAGGTC----- 744

Qy     241 uGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGl 261
                           :::||||||          ||||||
Db     745 ---------------AGAGCTGCCTCAGCCCCTGCCAAGGGG-----------ACCCC 776

Qy     261 nGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrTh 281
               |||      |||   :::|||         :::         |||   |||:::|||
Db     777 TGGGAAAGGGGCTACCCCAGCACCCCCTGGGAAGGCAGGGGCTGTAGCCTCCCAGACCAA 836

Qy     281 rSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPh 301
            :::
Db     837 GGCAGGG------------------------------------------------------ 843

Qy     301 eLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGl 321
                                                    :::   |||::::::::||
Db     844 ----------------------------------------AAGCCAGAGGAGGACTCAGA 863

Qy     321 uLysValAspGluGluLysGluLysThrGluProAlaSerGluGluGlnGluPro-- 340
           |         ||||||     :::               |||||||||:::   |||
Db     864 GAGCAGCAGCGAGGAGTCATCTGAC--------------AGTGAGGAGGAGACGCCAGC 908

Qy     341 -AlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLe 360
            |||:::         ||||||:::   |||
Db     909 TGCCAAGGCCCTGCTTCAGGCGAAGGCCTCAGGAAAA----------------------- 945

Qy     360 uGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu------------------- 373
                 |||||||||         |||   ::::::|||
Db     946 -ACCTCTCAGGTCGGAGCTGCCTCAGCCCCTGCCAAGGAGTCCCCCAGGAAAGGAGCTGC 1004

Qy     374 ---------------LysCysAlaProLeuAlaThrGluValPheAspGluLysMetGl 388
                          |||      |||         :::         |||   ||
Db    1005 CCCAGCGCCCCCTGGGAAGACAGGGCCTGCAGTTGCCAAGGCCCAGGCGGGGAAGCGGGA 1064

Qy     388 uAlaHisGlnGluValValAlaGluValHisValSerThrValGluLysThrGluGluGl 408
                     :::       :::|||                  |||      ||||||
Db    1065 GGAGGACTCGCAGAGCAGCAGCGAG-----------------GAATCGGACAGTGAGGA 1106

Qy     408 uGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThrGlyGluSerLe 428
           |:::                     |||      :::                  |||
Db    1107 GGAGGCGCCTGCTCAGGCGAAGCCTTCAGGGAAGGCCCCCAGGTCAGAGCCGCCTCGGC 1166

Qy     428 uPro------------ProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGl 444
            |||            |||   |||   |||   |||         |||::::::
Db    1167 CCCTGCCAAGGAGTCCCCCAGGAAAGGGGCTGCCCCAGCACCT---CCTAGGAAAACAGG 1223

Qy     444 uProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGl 464
            ||||||                              |||      |||         ::
Db    1224 GCCTGCAGCCGCCCAG-------------------GTCCAGGTGGGGAAGCAGGAGGA 1262

Qy     464 nLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleValSe 484
           :   :::      |||   ||||||                                ||
Db    1263 GGACTCAAGAAGCAGCAGCGAGGAGTCAGAC-------------------------AG 1295

Qy     484 rGluValGluMetLeuSerSerGlnGluArgIleLysValGln---------GlySerPr 501
           |:::   |||   |||::::::         :::|||:::            |||||
Db    1296 TGACAGAGAAGCACTGGCAGCCATGAATGCAGCTCAGGTGAAGCCCTTGGGGAAAAGCCC 1355

Qy     501 oLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGl 521
            |      |||      |||:::   |||:::      |||                 ||
Db    1356 CCAGGTGAAACCTGCCTCTACCATGGGCATGGGGCCCTTG-----------------GG 1397

Qy     521 yLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSe 541
            ||||      |||   |||      ||||||:::
Db    1398 GAAA------GGCGCCGGCCCAGTGCCACCTGGGAAGGTG-----------------GG 1433

Qy     541 rProGluSerAlaAspGluGlnLysGly---------GluSerSerAlaSerSerProGl 558
            |||   :::         |||   |||         |||   |||   ||||||   ||
Db    1434 GCCTGCAACCCCTCAGCCCAGGTGGGGAAGTGGGAGGAGGACTCAGAGAGCAGTAGTGA 1493

Qy     558 uGluProGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGl 578
            ||||      :::|||      :::|||:::   |||      |||:::         ::
Db    1494 GGAGT--CATCAGACAGCAGTGATGGAGAGGTGCCCACAGCTGTGGCCCCGGCTCAGGAA 1551

Qy     578 uAlaGluGluGlyThrThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAl 598
            ::::      ||||||:::|||                  |||         |||
```

```
Db   1552 AAGTCCTTGGGGAACATCCTCC---------------AGGCCAAACCCACCTCCA----- 1591
Qy    598 aSerPheLysLysMetValThrProLysLysArgValArgArg------------ProSe 614
Db   1592 ----------------GTCCTGCCAAGGGGCCCCCTCAGAAGGCAGGGCCTGTAGCCGTC 1635
Qy    614 rGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeuSerSerThrAs 634
Db   1636 CAGGTCAAG-GCTGAAAAGCCCATGGACAACTCGGAGAGCAGCGAG---GAGTCGTCGGA 1691
Qy    634 pSerThrValSerGluMetGlnAspGluValLysThrValGlyGlu-------------- 649
Db   1692 CAGTGCGGACAGTGAGGAGGCACCAGCAGCCATGACTGCAGCTCAGGCAAAACCAGCTCT 1751
Qy    650 -------GluGlnLysProGluGluProLysArgArgValAspThrSerValSerTrpGl 667
Db   1752 GAAAATTCCTCAGACCAAGGCCTGCCCAAAGAAAACCAATACCACTGCATCTGCCAAGGT 1811
Qy    667 uAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAs 687
Db   1812 CGCCCCTGTGCGAGTGGGCACCCAACCCCCCGGAAAGCAGGAACTGCGACTTCTCCAGC 1871
Qy    687 pGluGlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAs 707
Db   1872 AGGCTCATCCCCAGCTGTGGCTGGGGGCACCCAGAGACCAGCAGAGGATTCTTCAAGCAG 1931
Qy    707 pLysGluAlaGlyThrAspAlaValPro---------AlaSerThrGlnGluGlnAspGl 724
Db   1932 TGAGGAATCAGATAGTGAGGAAGAGAAGACAGGTCTTGCAGTAACCGTGGGACAGGCAAA 1991
Qy    724 nAlaGlnGlySerSerSerProGluProAlaGlySer------------ProSerGluGl 740
Db   1992 GTCTGTGGGGAAAGGCCTCCAGGTGAAAGCAGCCTCAGTGCCTGTCAAGGGGTCCTTGGG 2051
Qy    740 yGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLys-------- 757
Db   2052 GCAAGGGACTGCTCCA--------------------GTACTCCCTGGGAAGACGGGGCC 2090
Qy    758 ----------LysSerLysSerLysLeuGluGluLysAlaGluAspSerSerValGluGl 774
Db   2091 TACAGTCACCCAGGTGAAAGCTGAAAAGCAGGAAGACTCTGAGAGCAGTGAGGAGGAATC 2150
Qy    774 nLeuSerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIl 794
Db   2151 AGACAGTGAGGAAGCAGCTGCATCTCCAGCACAGGTGAAAACCTCAGTAAAGAAAACCCA 2210
Qy    794 eProGlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSe 814
Db   2211 GGCCAAAGCCAACCCAGCTGCCGCCAGAGCACCTTCAGCAAAAGGGACAATT-------- 2262
Qy    814 rGlyProValGluIleAsnGluAspAspProAsnValProAlaValValProLeuSerGl 834
Db   2263 ------------------------------TCAGCCCCTGGAAAAGTTGTCACTGCAGC 2291
Qy    834 uTyrAsnAlaValGluArgGluLysMetGluAlaGlnGlyAsnThrGluLeuProGlnLe 854
Db   2292 TGCTCAAGCCAAGCAGAGGTCTCCATCCAAGGTGAAGCCACCAGTGAGAAACCCCCAG-- 2349
Qy    854 uLeuGlyAlaValTyrValSerGluGluLeuSerLys------ThrLeuValHisThrVa 872
Db   2350 ----------------AACAGTACCGTCTTGGCGAGGGGCCCAGCATCTGTGCCATCTGT 2393
Qy    872 lSerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSerProSe 892
Db   2394 GGGGAAGGCCGTGGCTACAGCAGCTCAGGCCCAGACAGGGCCAGAGGAGGACTCAGGGAG 2453
Qy    892 rTrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetPro-- 911
Db   2454 CAGTGAGGAGGAGTCAGACAGTGAGGAGGAGGCGGAGACGCTGGCTCAGGCGAAGCCTTC 2513
Qy    912 ----------------------------------ProValGluGluValThrGluLysAs 920
Db   2514 AGGGAAGACCCACCAGATCAGAGCTGCCTTGGCTCCTGCCAAGGAGTCCCCCAGGAAAGG 2573

Qy    920 pIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAla-- 939
Db   2574 G------GCTGCCCCAACACCTCCT---GGGAAGACAGGGCCTTCGGCTGCCCAGGCAGG 2624
Qy    940 ----HisAspAspMetValThrSer------GluValAspPheThrSerGluAlaValTh 956
Db   2625 GAAGCAGGATGACTCAGGGAGCAGCAGCGAGGAATCAGACAGTGATGGGGAGGCACCGGC 2684
Qy    956 rAlaThrGluThrSerGluAlaLeuArg-------------------------------- 965
Db   2685 AGCTGTGACCTCTGCCCAGGTGATTAAACCCCCCCTGATTTTTGTCGACCCTAATCGTAG 2744
Qy    965 ------------------------------------------------------------ 965
Db   2745 TCCAGCTGGCCCAGCTGCTACACCCGCACAAGCCCAGGCTGCAAGCACCCCGAGGAAGGC 2804
Qy    966 -------ThrGluGluValThrGluAlaSerGlyAlaGlu---GluThrThrAspMetVa 982
Db   2805 CCGAGCCTCGGAGAGCACAGCCAGGAGCTCCTCCTCCGAGAGCGAGGATGAGGACGTGAT 2864
Qy    982 lSerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGl 1002
Db   2865 CCCCGCTACACAATGCTTGACTCCTGGCATCAGAACCAATGTGGTGACCATGCCCACTGC 2924
Qy   1002 nGluValGlu--------------------SerGlyValLeuAspThrGluGluGluGl 1015
Db   2925 CCACCCAAGAATAGCCCCCAAAGCCAGCATGGCTGGGGCCAGCAGCAGCAAGGAGTCCAG 2984
Qy   1015 uArg-------------------------------------------------------- 1016
Db   2985 TCGGATATCAGATGGCAAGAAACAGGAGGGACCAGCCACTCAGGTGTCAAAGAAGAACCC 3044
Qy   1017 ------------GlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGlu-- 1031
Db   3045 AGCTTCCCTCCCACTGACCCAGGCTGCCCTGAAGGTCCTCGCCCAGAAAGCCAGTGAGGC 3104
Qy   1032 ----------------------------------------GluSerGlnVal-------- 1035
Db   3105 TCAGCCTCCTGTTGCCAGGACCCAGCCTTCAAGTGGGGTTGACAGTGCTGTGGGAACACT 3164
Qy   1036 -------------ProAlaThrGlnThrValGlnArgThrGly----------------- 1045
Db   3165 CCCTGCAACAAGTCCCCAGAGCACCTCCGTCCAGGCCAAAGGGACCAACAAGCTCAGAAA 3224
Qy   1046 ----------------------------SerLysAlaLeuGluLysValGluGluValGl 1056
Db   3225 ACCTAAGCTTCCTGAGGTCCAGCAGGCCACCAAAGCCCCTGAGAGCTCAGATGACAGTGA 3284
Qy   1056 uGluAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyProVa 1076
Db   3285 GGACAGCAGCGACAGTTCTTCAGGGAGTGAGGAAGATGGTGAAGGGCCCCAGGGG----- 3339
Qy   1076 lGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGl 1096
Db   3340 -------------------------GCCAAGTCAGCCCACACGCTGGGTCCCACCCCCTC 3374
Qy   1096 uSerLeuGlu------ValProGluValThrAlaAspVal---------AspHisValAl 1111
Db   3375 CAGGACAGAGACCCTGGTGGAGGAGACCGCAGCAGAGTCCAGCGAGGATGATGTGGTGGC 3434
Qy   1111 aThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnAlaValAlaProGluSerSe 1131
Db   3435 GCCATCCCAGTCTCTC---CTCTCAGGTTATATGACCCCTGGACTAACCCCAGCCAATTC 3491
Qy   1131 rGluThr---------------LeuThrAspSerGluThrAsnGlySerThrProLeuAl 1146
Db   3492 CCAGGCCTCAAAAGCCACTCCCAAGCTAGATTCCAGCCCCTCAGTTTCCTCTACTCTGGC 3551
Qy   1146 aAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGlnAspSerLy 1166
Db   3552 CGCCAAAGATGACCCAGATGGCAAGCAGGAGGCAAAG------------------CCCCA 3593
```

```
Qy   1166 sAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGluAlaAlaThrAlaGlnLy 1186
          :     |||   :::          :::          :::|||||||||:::
Db   3594 ACAGGCAGCAGGCATGTTGTCCCCTAAAACAGGTGGAAAAGAGGCTGCTTCA-------- 3645

Qy   1186 sGluGluProSerThrLeuProAsnAsnValProAlaGlnGluGluHisGlyGluGluPr 1206
                   |||   |||      ||||||     :::|||   |||    |||||
Db   3646 ---------GGCACCACACCTCA-GAAGTCCCGGAAGCCCAAGAAAGGGGCTGGGAACC 3694

Qy   1206 oGlyArgAspValLeuGluPro------ThrGlnGlnGluLeuThrAlaAlaAlaValPr 1224
          |          |||  |||                                     ||
Db   3695 C------CCAAGCCTCAACCCTGGCGCTGCAAAGCAACATCACCCAGTGCCTCCTGGGCC 3748

Qy   1224 oValLeuAlaLys-ThrGluValGlyGlnGluGlyGluValAspTrpLeuAspGlyGluL 1244
          |  ||||||         |||        :::      |||      :::       |||
Db   3749 AACCCTGGCCCCTGAATGAGGCCCAGGTGCAGGCCTCAGTGGTGAAGGTCCTGACTGAGC 3808

Qy   1244 ysValLysGluGluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaA 1264
            :::::::::|||::::::            |||   :::         |||   :::
Db   3809 TGCTGGAACAGGAAAGAAAGAAGGTGGTGGACACCACCAAGGAGAGCAGCAGGAAGGGCT 3868

Qy   1264 spValThrTyrAspSerGluValMetGly---ValAlaGlyCysGlnGluLysGluSerT 1283
               :::      :::|||   :::|||         |||    ||||||:::   :::
Db   3869 GGGAGAGCCG-CAAGCGGAAGCTATCGGGAGACCAGCCAGCTGCCAGGACCCCCAGGAGC 3927

Qy   1283 hrGluValGln-SerLeuSerLeu---GluGluGlyGluMetGluThrAspValGluLys 1301
            |||   :::     |||          |||  ||||||                ||||||
Db   3928 AAGAAGAAGAAGAAGCTGGGGGCCGGGGAAGGTGGGGAGGCCTCTGTTTCCCCAGAAAAG 3987

Qy   1302 GluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAla 1321
                      :::|||   :::         ::::::        :::
Db   3988 ACCTCCACGACTTCCAAGGGGAAAGCAAAGAGAGACAAAGCAAGTGGTGATGTCAAGGAG 4047

Qy   1322 ProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGlu 1341
             :::   :::|||:::   |||                 |||   |||       |||
Db   4048 AAGAAAGGGAAGGGGTCTCTTGGCTCCCAAGGG---GCCAAGGACGAGCCAGAAGAGGAG 4104

Qy   1342 ArgGlyLysAlaLeuGlySerLeu---GlyGlySerProSerLeuProAspGlnAspLys 1360
                |||   :::|||::::::    ||||||      |||   |||          |||
Db   4105 CTTCAGAAGGGGATGGGGACGGTTGAAGGTGGAGATCAAAGCAACCCAAAGAGCAAGAAG 4164

Qy   1361 AlaGlyCysIleGluValGlnValGlnSerLeuAspThrThrValThrGlnThrAlaGlu 1380
                                  :::   :::   |||                  :::
Db   4165 -----------------GAGAAGAAGAAATCCGACAAGAGAAAAAAAGACAAAGAAAAA 4206

Qy   1381 AlaValGluLysValIleGluThrValValIleSerGluThrGlyGluSerProGluCys 1400
                :::|||                   ::::::      |||||||||
Db   4207 AAAGAAAAGAAGAAGAAAGCAAAAAAGGCCTCAACCAAAGATTCTGAGTCACCG------ 4260

Qy   1401 ValGlyAlaHisLeuLeuProAlaGluLysSerSerAla--------------------- 1413
                                     :::|||   :::|||
Db   4261 ---TCCCAGAAGAAAAAGAAGAAAAAGAAGAAGACAGCAGAGCAGACTGTATGACGAGCA 4317

Qy   1414 ------------------------------ThrGlyGlyHisTrpThrLeuGlnHisAla 1423
                                         :::   ||||||            ||||||
Db   4318 CCAGCACCAGGCACAGGGATTTCCTAGCCGAGCAGTGGCCAT---------CCCCATGCC 4368

Qy   1424 GluAsp------------------------------------------ThrValProLeu 1429
             |||                                          |||   |||
Db   4369 TCTGACCTCCACCGACCTCTGCCCACCATGGGTTGGAACTAAACTGTTACCTTCCCTCGC 4428

Qy   1430 GlyProGlu------SerGlnAlaGluSerIleProIleIleValThrProAlaProGlu 1447
             |||          ||||||   :::    :::|||:::::::    |||
Db   4429 TCCACAGAAGAAGACAGCCAGCTTCAGGG-GTCCCTGTGCTGGCCAAGCCAGTG------ 4481

Qy   1448 SerThrLeuHisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGlu 1467
                            |||:::|||      |||          :::|||:::    :::
Db   4482 ---------------AGCCTGCGGGGAGGCTGGTCC---------AAGGAGAAAGTGGAC 4517

Qy   1468 GluGluAspLysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGlu 1487
          :::         |||         |||                           ::::::
Db   4518 CAGCTCCCATGACCTCACCCCACTCCCCAACACAGGACGCTTCATATAGATGTGTACAG 4577

Qy   1488 LysValLeuLysAlaGlu---ProGluIleLeuGluLeuGluSerLysSerAsnLysIle 1506
             :::          :::     |||    :::|||
Db   4578 TATATGTATTTTTTAAGTGACCTCCTCTCCTT-------------------------- 4610

Qy   1507 ValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaProGlu 1526
                                                                |||:::
Db   4611 --------------------------------------------------------CCACAG 4616

Qy   1527 Thr---HisAlaTyrAspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGlu 1545
          |||      ||||||            :::              |||     |||       |||
Db   4617 ACCCCACATGCCCAAAGGCCTCGGGACTTCCCACCACCTTGCTCCACAGATCCAGCTAGG 4676

Qy   1546 ProAsnArgCysTrpThrLysMetLysAspAlaLysMetLysHisProValProGlnPro 1565
          |||:::   |||                                  ||||||||||||
Db   4677 CCTGACCTGTGCCT----------------------------CATCCCGTGCCG------ 4702

Qy   1566 ArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLeu 1585
                                       |||      |||      ||||||    ||||||
Db   4703 ---------------------------CTCGGTCTCTGGCTGATCCCGAGGCTTTGTCTT 4735

Qy   1586 Pro 1586
          |||
Db   4736 CCT 4738
```

RESULT 15
US-09-880-107-3311
; Sequence 3311, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3311
;   LENGTH: 11185
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: Genbank Accession No. US20020142981A1 U16306
US-09-880-107-3311

```
Alignment Scores:
Pred. No.:              1.98e-18    Length:          11185
Score:                  395.00      Matches:         431
Percent Similarity:     31.56%      Conservative:    252
Best Local Similarity:  19.92%      Mismatches:      725
Query Match:            4.89%       Indels:          756
DB:                     10          Gaps:            102

US-09-902-432-4 (1-1596) x US-09-880-107-3311 (1-11185)

Qy     25 ValLeuSerGlyHisGlyPro---AlaAlaGluAlaSerGlyAlaAlaGlyAspProAla 43
          ||||||      ||||||         :::   ||||||:::         :::
Db   2622 GTGCTTTTGGCCCATGGTACTTTAAGTGTTGAAGCAGCCACTGTATCAAAATGGTCATGG 2681

Qy     44 AspAlaAspProAlaThr---------------------------------------Lys 50
          |||   |||       |||                                       |||
Db   2682 GATGAAGATAATACAACATCCAAGCCTTTAGAGTCTACAGAACCTTCAGCCTCTTCAAAA 2741

Qy     51 LeuProGln--------------------LysAsnGlyGlnLeuSerSerValAsnGly 63
          ||||||                         ||||||:::          :::
Db   2742 TTGCCCCTGCCTTACTCACAACTGTGGGATGAATGGAAAGGATAAAGACATCCCAAGT 2801

Qy     64 ValAlaGluGlnGly------AspValHisValGlnGluGluAsnGlnGluGlyGlnGlu 81
```

```
Db  2802 TTCACTGAAGATGGAGCAGATGAATTTACTCTTATTCCAGATAGTACTCAAAAGCAGTTA 2861

Qy    82 GluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAsp 101
Db  2862 GAGGAGGTTACTGATGAAGACATAGCAGCCCATGGAAAATTCACAATTAGATTTCAGCCA 2921

Qy   102 ArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGln 121
Db  2922 ACTACATCAACTGGTATTGCAGAAAAGTCAACTTTGAGAGATTCTACA------------ 2969

Qy   122 GluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnVal---GluGlu 140
Db  2970 ------ACTGAAGAAAAAGTTCCACCTATCACAAGCACTGAAGGCCAAGTTTATGCAACC 3023

Qy   141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
Db  3024 ATGGAAGGAAGTGCTTTGGGTGAAGTAGAAGATGTGGACCTCTCTAAGCCAGTATCTACT 3083

Qy   161 Val------------------------------GlyPheLysPheThrValLysLysAsp 170
Db  3084 GTTCCCCAATTTGCACACACTTCAGAGGTGGAAGGATTAGCATTTGTTAGTTATAGTAGC 3143

Qy   171 LysAsnGlu------------------LysSerAspThrValGlnLeuLeuThrValLys 184
Db  3144 ACCCAAGAGCCTACTACTTATGTAGACTCTTCCCATACCATTCCTCTTTCTGTAATTCCC 3203

Qy   185 LysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis------GlnGlu 202
Db  3204 AAGACAGACTGGGGAGTGTTAGTACCTTCTGTTCCATCAGAAGATGAAGTTCTAGGTGAA 3263

Qy   203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSer------------Glu 218
Db  3264 CCCTCTCAAGACATACTTGTCATTGATCAGACTCGCCTTGAAGCGACTATTTCTCCAGAA 3323

Qy   219 LeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGlu------------Gln 234
Db  3324 ACTATGAGAACAACAAAAATCACAGAGGGAACAACTCAGGAAGAATTCCCTTGGAAAGAA 3383

Qy   235 SerSerThrGluIleProLeuGlnAla--------------------------------- 243
Db  3384 CAGACTGCAGAGAAACCAGTTCCTGCTCTCAGTTCTACAGCTTGGACTCCCAAGGAGGCA 3443

Qy   244 ---------------------GluSerAspGlnAlaAlaGluGluGluAlaLysAspGlu 256
Db  3444 GTAACACCACTGGATGAACAAGAGGGCGATGGATCAGCATATACAGTCTCTGAAGATGAA 3503

Qy   257 ---------GlyGluGlu------------------------------------------ 259
Db  3504 TTGTTGACAGGTTCTGAGAGGGTCCCAGTTTTAGAAACAACTCCAGTTGGAAAAATTGAT 3563

Qy   260 ---------------------------------------------LysGlnGluLysGlu 264
Db  3564 CACAGTGTGTCTTATCCACCAGGTGCTGTAACTGAGCACAAAGTGAAAACAGATGAAGTG 3623

Qy   265 ProThrLysSerProGlu---------------------SerProSerSerProValAsn 277
Db  3624 GTAACACTAACACCACGCATTGGCCAAAAGTATCTTTAAGTCCAGGGCCTGAACAAAAA 3683

Qy   278 SerGluThr------------------------ThrSerSerPheLysLysPhePheThr 289
Db  3684 TATGAAACAGAAGGTAGTAGTACAACAGGATTTACATCATCTTTGAGTCCTTTTAGTACC 3743

Qy   290 HisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAsp---Asp 308
Db  3744 CACATTACCCAGCTTATGGAAGAAACCACTACTGAGAAAACATCCCTAGAGGATATTGAT 3803

Qy   309 LeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLys 328
Db  3804 TTAGGCTCAGGATTATTTGAAAAGCCCAAAGCCACAGAACTCATAGAATTTTCAACAATC 3863

Qy   329 GluLysThrGluProAlaSer--------------------------------------- 335
Db  3864 AAAGTCACAGTTCCAAGTGATATTACCACTGCCTTCAGTTCAGTAGACAGACTTCACACA 3923

Qy   336 ------------------------------------------------GluGluGlnGlu 339
Db  3924 ACTTCAGCATTCAAGCCATCTTCCGCGATCACTAAGAAACCACCTCTCATCGACAGGGAA 3983

Qy   340 ProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuPro 359
Db  3984 CCTGGTGAAGAAACAACCAGTGACATGGTAATCATTGGAGAATCAACATCTCATGTTCCT 4043

Qy   360 ---------LeuGluAspGlnValGly---------------AspLeuGluAlaSerSer 371
Db  4044 CCCACTACCCTTGAAGATATTGTAGCCAAGGAAACAGAAACCGATATTGATAGAGAGTAT 4103

Qy   372 GluGluLysCysAlaProLeuAlaThrGlu------------------ValPheAspGlu 385
Db  4104 TTCACGACTTCAAGTCCTCCTGCTACACAGCCAACAAGACCACCCACTGTGGAAGACAAA 4163

Qy   386 LysMetGluAlaHisGlnGluValValAla------------------------------ 395
Db  4164 GAGGCCTTTGGACCTCAGGCGCTTTCTACGCCACAGCCCCCAGCAAGCACAAAATTTCAC 4223

Qy   396 ---GluValHisValSerThrValGluLysThrGluGluGluGlnGlyGlyGlyGlyGlu 414
Db  4224 CCTGACATTAATGTTTATATTATTGAGGTCAGAGAAAATAAGACAGGTCGAATGAGTGAT 4283

Qy   415 AlaGluGlyGlyValValValGluGlyThrGlyGluSerLeuProProGluLysLeuAla 434
Db  4284 ------------------TTGAGTGTAATTGGTCATCCAATAGATTCAGAATCT---AAA 4322

Qy   435 GluProGlnGluValProGlnGluAlaGluProAlaGluGluLeuMet------------ 450
Db  4323 GAAGATGAACCTTGTAGTGAAGAAACAGATCCAGTGCATGATCTAATGGCTGAAATTTTA 4382

Qy   451 ------------------------------------------------------LysSer 452
Db  4383 CCTGAATTCCCTGACATAATTGAAATAGACCTATACCACAGTGAAGAAAATGAAGAAGAA 4442

Qy   453 ArgGluMetCysValSerGlyGlyAspHisThr---------------GlnLeuThrAsp 467
Db  4443 GAAGAAGAGTGTGCAAATGCTACTGATGTGACAACCACCCCATCTGTGCAGTACATAAAT 4502

Qy   468 LeuSerProGluGluLysThrLeuProLysHisProGluGly------------------ 481
Db  4503 GGGAAGCATCTCGTTACCACTGTGCCCAAGGACCCAGAAGCTGCAGAAGCTAGGCGTGGC 4562

Qy   481 ------------------------------------------------------------ 481
Db  4563 CAGTTTGAAAGTGTTGCACCTTCTCAGAATTTCTCGGACAGCTCTGAAAGTGATACTCAT 4622

Qy   482 ------IleValSerGluValGluMet---------------------------LeuSer 490
Db  4623 CCATTTGTAATAGCCAAAACGGAATTGTCTACTGCTGTGCAACCTAATGAATCTACAGAA 4682

Qy   491 SerGlnGluArgIleLysValGlnGlySer------------------------------ 500
Db  4683 ACAACTGAGTCTCTTGAAGTTACATGGAAGCCTGAGACTTACCCTGAAACATCAGAACAT 4742

Qy   501 ------------------------------------ProLeuLysLysLeuPheSerSer 508
Db  4743 TTTTCAGGTGGTGAGCCTGATGTTTTCCCCACAGTCCCATTCCATGAGGAATTTGAAAGT 4802

Qy   509 SerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArgGlyGlyGlyGlyAsp 528
Db  4803 GGAACAGCCAAAAAA------GGGGCAGAATCAGTCACAGAGAGA------------GAT 4844

Qy   529 GluGluProGlyGluTyrGlnHisIleHisThrGlu------------------------ 540
Db  4845 ACTGAAGTTGGTCATCAGGCACATGAACATACTGAACCTGTATCTCTGTTTCCTGAAGAG 4904

Qy   541 ---SerProGluSerAlaAspGluGlnLysGlyGluSerSerAla--------------- 554
Db  4905 TCTTCAGGAGAGATTGCCATTGACCAAGAATCTCAGAAAATAGCCTTTGCAAGGGCTACA 4964
```

```
Qy    555 ------SerSerProGluGluProGluGluThrThrCysLeuGluLysGlyProLeuGlu 572
                :::        |||||    |||::::::|||    :::           |||
Db   4965 GAAGTAACATTTGGTGAAGAGGTAGAAAAAAGTACTTCTGTCACATACACTCCCACTATA 5024

Qy    573 AlaProGlnAspGlyGluAla------------GluGluGlyThrThrSerAspGlyGlu 588
             |||          |||             ||||||     |||       |||
Db   5025 GTTCCAAGTTCTGCATCAGCATATGTTTCAGAGGAAGAAGCAGTTACCCTAATAGGA--- 5081

Qy    589 LysLysArgGluGlyIleThrProTrp------------AlaSerPheLysLysMet--- 603
                                 |||||            |||    |||:::
Db   5082 ------------------AATCCTTGGCCAGATGACCTGTTGTCTACCAAAGAAAGCTGG 5123

Qy    604 ------ValThrProLysLysArgVal--------------------------------- 610
                   ||||||::::::   |||
Db   5124 GTAGAAGCAACTCCTAGACAAGTTGTAGAGCTCTCAGGGAGTTCTTCGATTCCAATTACA 5183

Qy    611 ArgArgProSerGluSerAspLysGluGluGluLeuGluLysValLysSerAlaThrLeu 630
                      |||:::::::::::::|||:::                      |||
Db   5184 GAAGGCTCTGGAGAAGCAGAAGAAGATGAAGATACAATGTTCACCATGGTAACTGATTTA 5243

Qy    631 SerSerThrAspSerThr---------------------------------ValSerGlu 639
          |||      ::::::|||                                 ::::::|||
Db   5244 TCACAGAGAAATACTACTGATACACTCATTACTTTAGACACTAGCAGGATAATCACAGAA 5303

Qy    640 MetGlnAspGluValLysThrValGlyGluGluGlnLysProGluGluProLysArgArg 659
                   ||||||                            |||:::|||      :::
Db   5304 AGCTTTTTGAGGTTCCTGCAACCACCATTTATCCAGTTTCTGAACAACCTTCTGCAAAA 5363

Qy    660 Val---------------------------AspThrSerValSerTrpGluAlaLeuIle 670
          |||                           |||||||||      |||    :::
Db   5364 GTGGTGCCTACCAAGTTTGTAAGTGAAACAGACACTTCT---GAGTGGATTTCCAGTACC 5420

Qy    671 CysValGlySerSerLysLysArgAlaArgLysAlaSerSerSerAspAspGluGlyGly 690
             |||         |||::::::                     ::::::||||||
Db   5421 ACTGTTGAGGAAAAGAAAAGGAAG---------------------GAGGAGGAGGGAACT 5459

Qy    691 ProArgThrLeuGlyGlyAspSerHisArgAlaGluGluAlaSerLysAspLysGluAla 710
                |||                                                |||
Db   5460 ACAGGTACG------------------------------------------------GCT 5471

Qy    711 GlyThrAspAlaValProAlaSerThrGlnGluGlnAspGln---------AlaGlnGly 727
             |||       |||   :::|||||||||      ||||||
Db   5472 TCTACATTTGAGGTATATTCATCTACACAGAGATCGGATCAATTAATTTTACCCTTTGAA 5531

Qy    728 SerSerSerProGluProAlaGlySerProSerGluGlyGluGlyValSerThrTrpGlu 747
                ||||||      |||    |||            |||         |||   :::
Db   5532 TTAGAAAGTCCAAATGTAGCTACATCTAGTGATTCAGGT------------ACCAGGAAA 5579

Qy    748 SerPheLysArgLeuValThrProArgLysLysSerLysSerLysLeuGluGluLys--- 766
          ||||||      |||   ||||||   :::         :::         :::
Db   5580 AGTTTTATGTCCTTGACAACACCAACACAGTCTGAAAGGGAAATGACAGATTCTACTCCT 5639

Qy    767 ------AlaGluAspSerSerValGluGlnLeuSerThrGluIleGluProSerArgGlu 784
                   |||   :::::::::|||   |||            :::      |||
Db   5640 GTCTTTACAGAAACAAATACATTAGAAAATTTGGGGGCACAG---------ACCACTGAG 5690

Qy    785 GluSerTrpValSerIleLysLysPheIleProGlyArgArgLysLysArgAlaAspGly 804
             |||   :::                                          |||
Db   5691 CACAGCAGTATC------------------------------------CATCAACCTGGG 5714

Qy    805 LysGlnGluGlnAlaThrValGluAspSerGlyProValGluIleAsnGluAsp------ 822
             ||||||   |||                  |||      :::      :::
Db   5715 GTTCAGGAAGGGCTGACCACTCTCCCACGTAGTCCTGCCTCTGTCTTTATGGAGCAGGGC 5774

Qy    823 ------------------AspProAsnValProAlaValValProLeuSer--------- 833
                            ||||||            |||      |||
Db   5775 TCTGGAGAAGCTGCTGCCGACCCAGAAACCACCACTGTTTCTTCATTTTCATTAAACGTA 5834

Qy    834 GluTyrAsnAlaValGluArgGluLys---------------------------MetGlu 844
          ||||||   |||::::::   ||||||                           :::|||
Db   5835 GAGTAT---GCAATTCAAGCCGAAAAGGAAGTAGCTGGCACTTTGTCTCCGCATGTGGAA 5891

Qy    845 AlaGlnGlyAsnThrGlu------------------------------------------ 850
             :::||||||
Db   5892 ACTACATTCTCCACTGAGCCAACAGGACTGGTTTTGAGTACAGTAATGGACAGAGTAGTT 5951

Qy    851 ---------LeuProGlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLys--- 866
                   :::   |||            :::   :::||||||   |||   :::
Db   5952 GCTGAAAATATAACCCAAACATCCAGGGAAATAGTGATTTCAGAGCGATTAGGAGAACCA 6011

Qy    866 ------------------------------------------------------------ 866

Db   6012 AATTATGGGGCAGAAATAAGGGGCTTTTCCACAGGTTTTCCTTTGGAGGAAGATTTCAGT 6071

Qy    867 ---------------------ThrLeuValHisThrValSer---------ValAlaVal 876
                               |||:::   |||   ::::::               :::
Db   6072 GGTGACTTTAGAGAATACTCAACAGTGTCTCATCCCATAGCAAAAGAAGAAACGGTAATG 6131

Qy    877 IleAspGlyThr---ArgAlaValThrSerValGluGluArgSerProSerTrpIleSer 895
          ::::::|||:::      |||                  :::   |||||||||   :::
Db   6132 ATGGAAGGCTCTGGAGATGCAGCATTTAGGGACACCCAGACTTCACCATCTACAGTACCT 6191

Qy    896 AlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetPro------------ 911
          ||||||         :::   |||   :::         |||
Db   6192 ACTTCAGTT------CACATCAGTCACATATCTGACTCAGAAGGACCCAGTAGCACCATG 6245

Qy    912 ------------------ProValGluGluValThrGlu------------------Lys 919
                            |||   ||||||   |||                      :::
Db   6246 GTCAGCACTTCAGCCTTCCCCTGGGAAGAGTTTACATCCTCAGCTGAGGGCTCAGGTGAG 6305

Qy    920 AspIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAla 939
             ::::::      :::   |||:::         ||||||
Db   6306 CAACTGGTCACAGTCAGCAGCTCTGTTGTT---CCAGTGCTTCCCAGTGCTGTGCAAAAG 6362

Qy    940 HisAspAspMetValThrSerGluValAsp---PheThrSerGluAlaValThrAlaThr 958
                   :::|||   :::|||               |||      |||
Db   6363 TTTTCTGGTACAGCTTCCTCCATTATCGACGAAGGATTGGGAGAAGTGGGTACTGTCAAT 6422

Qy    959 Glu---------ThrSerGluAlaLeuArgThrGluGluVal------ThrGluAlaSer 973
          |||         |||   |||   |||   ||||||         |||:::|||
Db   6423 GAAATTGATAGAAGATCCACCATTTTACCAACAGCAGAAGTGGAAGGTACGAAAGCTCCA 6482

Qy    974 GlyAlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAsp 993
             :::|||         :::         ||||||      |||:::   |||
Db   6483 GTAGAGAAGGAGGAAGTAAAGGTCAGTGGCACAGTTTCA------ACAAACTTTCCCCAA 6536

Qy    994 ThrThrGlu------------------------GluAlaThrProVal---GlnGluVal 1004
          |||   |||                        |||      ||||||   ||||||:::
Db   6537 ACTATAGAGCCAGCCAAATTATGGTCTAGGCAAGAAGTCAACCCTGTAAGACAAGAAATT 6596

Qy   1005 GluSerGlyValLeuAspThrGluGluGluArgGlnThrGlnAlaIleLeuGlnAla 1024
          ||||||            |||   ||||||:::   |||      :::         ::::::
Db   6597 GAAAGTGAAACA------ACATCAGAGGAACAAATTCAAGAAGAAAAGTCATTTGAATCC 6650

Qy   1025 ValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGlnArgThr 1044
                :::            |||      :::      :::||||||            |||
Db   6651 CCTCAAAACTCTCCTGCAACAGAACAAACAATCTTTGATTCACAGACATTTACTGAAACT 6710

Qy   1045 GlySerLys--------------------------------------------------- 1047
             |||
Db   6711 GAACTCAAAACCACAGATTATTCTGTACTAACAACAAAGAAAACTTACAGTGATGATAAA 6770

Qy   1047 ------------------------------------------------------------ 1047

Db   6771 GAAATGAAGGAGGAAGACACTTCTTTAGTTAACATGTCTACTCCAGATCCAGATGCAAAT 6830

Qy   1048 ---------------------------AlaLeuGluLys--------------------- 1051
                                     |||   ||||||
Db   6831 GGCTTGGAATCTTACACAACTCTCCCTGAAGCTACTGAAAAGTCACATTTTTTCTTAGCT 6890

Qy   1052 ---------------------------ValGluGluValGluGluAspSerGluValLeu 1062
                                     |||   |||      ||||||   :::
Db   6891 ACTGCATTAGTAACTGAATCTATACCAGCTGAACATGTAGTCACAGATTCACCAATCAAA 6950

Qy   1063 AlaSerGluLysGluLysAspValMetProLysGly---------ProValGlnGluAla 1079
```

```
Db   6951 AAGGAAGAAAGTACAAAACAT---TTTCCGAAAGGCATGAGACCAACAATTCAAGAGTCA 7007

Qy   1080 GlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThr--------------- 1094

Db   7008 GATACTGAGCTCTTATTCTCTGGACTGGGATCAGGAGAAGAAGTTTTACCTACTCTACCA 7067

Qy   1095 ProGluSerLeuGluValProGluVal-----------------------------Thr 1104

Db   7068 ACAGAGTCAGTGAATTTTACTGAAGTGGAACAAATCAATAACACATTATATCCCCACACT 7127

Qy   1105 AlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeu------Met 1122

Db   7128 TCTCAAGTGGAAAGTACCTCAAGTGACAAAATTGAAGACTTTAACAGAATGGAAAATGTG 7187

Qy   1123 GluGlnAlaValAlaProGluSerSerGluThr-------------------------- 1133

Db   7188 GCAAAAGAAGTTGGACCACTCGTATCTCAAACAGACATCTTTGAAGGTAGTGGGTCAGTA 7247

Qy   1134 -----------------------LeuThrAspSerGluThrAsnGlySerThr------ 1143

Db   7248 ACCAGCACAACATTAATAGAAATTTTAAGTGACACTGGAGCAGAAGGACCCACGGTGGCA 7307

Qy   1144 ProLeuAlaAspSerAspThrAlaAspGlyThrGlnGlnAspGluThrIleAspSerGln 1163

Db   7308 CCTCTCCCTTTCTCC---ACGGACATCGGACATCCTCAAAATCAGACTGTCAGGTGGGCA 7364

Qy   1164 AspSerLysAlaThrAlaAlaValArgGlnSerGlnValThrGluGluGlu--------- 1180

Db   7365 GAAGAAATCCAGACTAGT------AGACCACAAACCATAACTGAACAAGACTCTAACAAG 7418

Qy   1181 ---AlaAlaThrAlaGlnLysGluGluProSerThrLeuProAsnAsnValProAlaGln 1199

Db   7419 AATTCTTCAACAGCAGAAATTAACGAAACAACAACCTCATCTACTGATTTTCTGGCTAGA 7478

Qy   1200 GluGluHisGlyGluGluProGlyArgAspValLeuGluProThrGlnGlnGluLeuThr 1219

Db   7479 ---GCTTATGGT--------------------TTTGAAATGGCCAAAGAATTTGTTACA 7514

Qy   1220 AlaAlaAlaValProValLeuAlaLysThrGluValGlyGlnGlu------GlyGluVal 1237

Db   7515 TCAGCACCAAAACCATCTGACTTGTATTATGAACCTTCTGGAGAAGGATCTGGAGAAGTG 7574

Qy   1238 AspTrpLeuAspGlyGluLysVal-----------------------LysGluGluGln 1249

Db   7575 GATATTGTTGATTCATTTCACACTTCTGCAACTACTCAGGCAACCAGACAAGAAAGCAGC 7634

Qy   1250 GluValPheValHisSerGly---------------ProAsnSerGlnLysAlaAlaAsp 1264

Db   7635 ACCACATTTGTTTCTGATGGGTCCCTGGAAAAACATCCTGAGGTGCCAAGCGCTAAAGCT 7694

Qy   1265 ValThrTyrAsp---------------SerGluValMetGlyValAlaGlyCysGlnGlu 1279

Db   7695 GTTACTGCTGATGGATTCCCAACAGTTTCAGTGATGCTGCCTCTTCATTCAGAGCAGAAC 7754

Qy   1280 LysGluSerThrGluValGlnSerLeu------SerLeuGluGluGlyGluMetGluThr 1297

Db   7755 AAAAGCTCCCCTGATCCAACTAGCACACTGTCAAATACAGTGTCATATGAGAGGTCCACA 7814

Qy   1298 AspValGluLysGluLysArgGluThrLysProGluGlnValSerGluGluGlyGluGln 1317

Db   7815 GACGGTAGTTTCCAAGACCGT--------------------TTCAGGGAATTCGAGGAT 7853

Qy   1318 GluThrAlaAlaProGluHisGluGlyThrTyrGlyLysProValLeuThrLeuAspMet 1337

Db   7854 TCCACCTTAAAACCTAACAGAAAAAAACCCACTGAAAATATTATCATAGACCTGGACAAA 7913

Qy   1338 ProSerSerGluArgGlyLysAlaLeuGlySerLeuGlyGlySerProSerLeuProAsp 1357

Db   7914 GAGGACAAGGATTTAATATTGACAATTACAGAGAGTACCATCCTTGAAATTCTACCTGAG 7973

Qy   1358 ------GlnAspLysAlaGlyCysIleGluVal-------------------------- 1366

Db   7974 CTGACATCGGATAAAAATACTATCATAGATATTGATCATACTAAACCTGTGTATGAAGAC 8033

Qy   1367 ---------GlnValGlnSerLeuAspThrThrVal----------------------- 1375

Db   8034 ATTCTTGGAATGCAAACAGATATAGATACAGAGGTACCATCAGAACCACATGACAGTAAT 8093

Qy   1376 --------------------ThrGlnThrAlaGluAlaValGluLysValIleGluThr 1388

Db   8094 GATGAAAGTAATGATGACAGCACTCAAGTTCAAGAGATCTATGAGGCAGCTGTCAACCTT 8153

Qy   1389 ValValIleSerGluThrGlyGlu---SerProGluCysValGlyAlaHisLeuLeuPro 1407

Db   8154 TCTTTAACTGAGGAAACATTTGAGGGCTCTGCTGATGTTCTGGCTAGCTACACTCAGGCA 8213

Qy   1408 AlaGluLysSerSerAlaThr----------------------------GlyGlyHis 1417

Db   8214 ACACATGATGAATCAATGACTTATGAAGATAGAAGCCAACTAGATCACATGGGCTTTCAC 8273

Qy   1418 TrpThrLeuGlnHisAlaGluAspThrValProLeuGlyProGluSerGlnAlaGlu--- 1436

Db   8274 TTCACA--------------ACTGGGATCCCT---GCTCCTAGCACAGAAACAGAATTA 8315

Qy   1437 -----------------------SerIleProIleIleValThrProAlaProGluSer 1448

Db   8316 GACGTTTTACTTCCCACGGCAACATCCCTGCCAATT---------CCTCGTAAGTCTGCC 8366

Qy   1449 ThrLeuHisProAspLeuGlnGly----------------------------------- 1456

Db   8367 ACAGTTATTCCAGAGATTGAAGGAATAAAAGCTGAAGCAAAAGCCCTGGATGACATGTTT 8426

Qy   1457 --------------------------------------------GluIle------Ser 1459

Db   8427 GAATCAAGCACTTTGTCTGATGGTCAAGCTATTGCAGACCAAAGTGAAATAATACCAACA 8486

Qy   1460 AlaSerGlnArgGluArgSerGluGlu------GluAspLysProAspAlaGlyProAsp 1477

Db   8487 TTGGGCCAATTTGAAAGGACTCAGGAGGAGTATGAAGACAAAAAACATGCTGGTCCTTCT 8546

Qy   1478 AlaAspGlyLysGluSerThrAlaIleGluLysValLeu-------------------- 1490

Db   8547 TTTCAGCCAGAATTCTCTTCAGGAGCTGAGGAGGCATTAGTAGACCATACTCCCTATCTA 8606

Qy   1491 --------------------------LysAlaGluProGluIleLeuGluLeuGluSer 1501

Db   8607 AGTATTGCTACTACCCACCTTATGGATCAGAGTGTAACAGAGGTGCCTGATGTGATGGAA 8666

Qy   1502 LysSerAsnLysIleValLeuAsnValIleGlnThrAlaValAspGlnPheAlaArg--- 1520

Db   8667 GGATCCAATCCCCCATATTACACTGATACAACATTAGCAGTTTCAACATTTGCGAAGTTG 8726

Qy   1521 ---------ThrGluThrAlaProGluThrHisAlaTyrAspSerGln----------- 1533

Db   8727 TCTTCTCAGACACCATCATCTCCCCTCACTATCTACTCAGGCAGTGAAGCCTCTGGACAC 8786

Qy   1534 ThrGlnValPro 1537

Db   8787 ACAGAGATCCCC 8798
```

Search completed: December 13, 2002, 06:51:00
Job time : 1298 secs

GenCore version 5.1.3


```
OM protein - nucleic search, using frame_plus_p2n model

Run on:         December 13, 2002, 00:22:56 ; Search time 4124 Seconds
                                          (without alignments)
                                          6267.700 Million cell updates/sec

Title:          US-09-902-432-4
Perfect score:  8073
Sequence:       1 MGAGSSTEQRSPEQPAGSDT..........AWAQPRKCLPRLQLKAPVSK 1596

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext  0.5
                Ygapop 10.0 , Ygapext  0.5
                Fgapop  6.0 , Fgapext  7.0
                Delop   6.0 , Delext   7.0

Searched:       16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters:      32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09902432/runat_12122002_141452_25812/app_query.fasta_1.1735
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09902432_@CGN_1_1_3512_@runat_12122002_141452_25812 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      EST:*
                1:  em_estba:*
                2:  em_esthum:*
                3:  em_estin:*
                4:  em_estmu:*
                5:  em_estov:*
                6:  em_estpl:*
                7:  em_estro:*
                8:  em_htc:*
                9:  gb_est1:*
                10:  gb_est2:*
                11:  gb_htc:*
                12:  gb_est3:*
                13:  gb_est4:*
                14:  gb_est5:*
                15:  em_estfun:*
                16:  em_estom:*
                17:  gb_gss:*
                18:  em_gss_hum:*
                19:  em_gss_inv:*
                20:  em_gss_pln:*
                21:  em_gss_vrt:*
                22:  em_gss_fun:*
                23:  em_gss_mam:*
                24:  em_gss_mus:*
                25:  em_gss_other:*
                26:  em_gss_pro:*
                27:  em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	1268.5	15.7	1269	11	BC015830	BC015830 Homo sapi
	2	1041.5	12.9	776	13	BG974884	BG974884 602842771
	3	1000	12.4	668	10	BB546856	BB546856 BB546856
	4	908	11.2	661	13	BI652823	BI652823 603298753
	5	856.5	10.6	674	9	AL045541	AL045541 DKFZp434K
	6	854.5	10.6	1117	14	BM804332	BM804332 AGENCOURT
	7	853.5	10.6	1153	14	D88456	D88456 D88456 Mous
	8	821	10.2	610	9	AI225639	AI225639 uj13g09.y
	9	806.5	10.0	646	10	BE618075	BE618075 601462263
c	10	785.5	9.7	796	9	AI796191	AI796191 wh43g05.x
	11	782	9.7	873	14	BQ717711	BQ717711 AGENCOURT
	12	776.5	9.6	924	14	BQ221404	BQ221404 AGENCOURT
	13	748	9.3	530	9	AI156819	AI156819 ui44c01.y
	14	744	9.2	679	14	BQ443215	BQ443215 UI-M-EV0-
c	15	713	8.8	573	12	BF400579	BF400579 UI-R-CA0-
c	16	701	8.7	584	12	BF448126	BF448126 hr82h12.x
c	17	690.5	8.6	532	10	BE350908	BE350908 ht63g05.x
	18	686.5	8.5	886	12	BF163549	BF163549 601769945
	19	686	8.5	641	10	BB665546	BB665546 BB665546
	20	682	8.4	875	14	BQ427917	BQ427917 AGENCOURT
	21	673.5	8.3	1045	12	BG394092	BG394092 602456258
	22	666.5	8.3	539	12	BG835245	BG835245 354322 MA
	23	657	8.1	644	13	BI446672	BI446672 dac81f11.
	24	656	8.1	490	9	AL041664	AL041664 DKFZp434M
	25	653	8.1	467	9	AL041405	AL041405 DKFZp434C
	26	637.5	7.9	911	12	BG751345	BG751345 602730292
	27	632	7.8	923	14	BQ919238	BQ919238 AGENCOURT
c	28	598.5	7.4	556	12	BG347898	BG347898 dac81f11.
	29	592.5	7.3	527	9	AL041657	AL041657 DKFZp434M
c	30	592	7.3	461	12	BF364956	BF364956 CM2-NN114
c	31	591.5	7.3	651	13	BJ044180	BJ044180 BJ044180
	32	590.5	7.3	861	13	BI154270	BI154270 602903096
	33	590	7.3	459	9	AL040086	AL040086 DKFZp434B
	34	577.5	7.2	466	12	BE859328	BE859328 UI-M-AQ0-
	35	568.5	7.0	724	14	BM721210	BM721210 UI-E-EO0-
	36	566.5	7.0	561	9	AL703803	AL703803 DKFZp686J
	37	560	6.9	812	12	BG538601	BG538601 602567330
c	38	557	6.9	465	12	BF364993	BF364993 CM2-NN114
	39	542	6.7	423	14	W82339	W82339 me98d02.r1
c	40	539	6.7	446	10	BE120002	BE120002 UI-R-CA0-
	41	537	6.7	431	9	AI322363	AI322363 me98d02.y
	42	513	6.4	1105	14	BQ935865	BQ935865 AGENCOURT
	43	506	6.3	883	13	BM047462	BM047462 603628755
	44	504	6.2	333	9	AI548287	AI548287 UI-R-C3-t
c	45	500.5	6.2	436	9	AA159505	AA159505 zo72c07.s

ALIGNMENTS

```
RESULT 1
BC015830
LOCUS       BC015830                1269 bp    mRNA    linear   HTC 17-OCT-2001
DEFINITION  Homo sapiens, clone IMAGE:4691862, mRNA.
ACCESSION   BC015830
VERSION     BC015830.1  GI:16198369
KEYWORDS    HTC.
SOURCE      Homo sapiens.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1269)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-OCT-2001) National Institutes of Health, Mammalian
```

```
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: CLONTECH
            cDNA Library Preparation: CLONTECH Laboratories, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
            Greene, Mark Ketteman and Anuradha Madan

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 38 Row: k Column: 11
            This clone was selected for full length sequencing because it
            passed the following selection criteria: matched mRNA gi: 4826749
            This clone has the following problem: frame shifted.
FEATURES             Location/Qualifiers
     source          1..1269
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4691862"
                     /tissue_type="Lung"
                     /clone_lib="NIH_MGC_77"
                     /lab_host="DH10B"
                     /note="Vector: pDNR-LIB"
BASE COUNT      419 a    311 c    365 g    174 t
ORIGIN

Alignment Scores:
Pred. No.:              2.64e-93      Length:        1269
Score:                  1268.50       Matches:       271
Percent Similarity:     75.06%        Conservative:  39
Best Local Similarity:  65.62%        Mismatches:    70
Query Match:            15.71%        Indels:        33
DB:                     11            Gaps:          8

US-09-902-432-4 (1-1596) x BC015830 (1-1269)

Qy        1 MetGlyAlaGlySerSerThrGluGlnArgSerProGluGln---ProAlaGlySerAsp 19
            ||||||||||||||||||||||||||||||||||||||||||   |||   |||||
Db       25 ATGGGCGCCGGGAGCTCCACCGAGCAGCGCAGCCCGGAGCAGCCGCCCGAGGGGAGCTCC 84

Qy       20 ThrProSerGluLeuValLeuSerGlyHisGlyProAlaAlaGluAlaSer---GlyAla 38
            |||||:::|||             ||||||:::|||||||||:::
Db       85 ACGCCGGCTGAGCCCGAGCCCAGCGGCGGCGGCCCCTCGGCCGAGGCGGCGCCAGACACC 144

Qy       39 AlaGlyAspProAlaAspAla------AspProAlaThrLysLeuProGlnLysAsnGly 56
                  ||||||||   |||      ||||||||||||||||   |||||||||||
Db      145 ACCGCGGACCCCGCCATCGCTGCCTCGGACCCCGCCACCAAGCTCCTACAGAAGAATGGT 204

Qy       57 GlnLeuSerSerValAsnGlyValAlaGluGlnGlyAspValHisValGlnGluGluAsn 76
            ||||||||::::::||||||||||||||||||   ::::::   :::||||||   :::
Db      205 CAGCTGTCCACCATCAATGGCGTAGCTGAGCAAGATGAGCTCAGCCTCCAGGAGGGTGAC 264

Qy       77 GlnGluGlyGln------------------------------------------GluGlu 82
                  ||||||                                          ||||||
Db      265 CTAAATGGCCAGAAAGGAGCCCTGAACGGTCAAGGAGCCCTAAACAGCCAGGAGGAAGAA 324

Qy       83 GluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLysAspArg 102
            ||||||:::     :::|||||||||||:::||||||||||||     ::::::|||
Db      325 GAAGTCATTGTCACAGAGGTTGGACAGAGAGACTCTGAAGATGTGAGCAAAAGAGACTCC 384

Qy      103 ValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGlyGlnGlu 122
               :::|||||||||      |||      |||   ||||||||   |||||||||||
Db      385 GATAAAGAGATGGCTACTAAGTCAGCGGTTGTTCACGACATCACAGATGATGGGCAGGAG 444

Qy      123 GluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetVal 142
            ||||||   ||||||||||||||||||||||:::||||||:::|||:::||||||:::
Db      445 GAGACACCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACA 504

Qy      143 GlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGly 162
            ||||||   |||||||||||||||||||:::||||||||||||||||||||||||||||||
Db      505 CAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGC 564

Qy      163 PheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThr 182
            ||||||||||||||||||||||||   ||||||   ||||||||||||||||||
Db      565 TTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACT 624

Qy      183 ValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGlu 202
            ||||||||||||||||||||||||                  |||||||||||||::::::
Db      625 GTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACAAGGAC 675

Qy      203 ProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSer 222
            ||||||:::         ||||||:::||||||||||||||||||   |||||||||
Db      676 CCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCT 729

Qy      223 ThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGln 242
            |||||||||   |||   |||||||||:::|||||||||         ||||||
Db      730 ACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCA 789

Qy      243 AlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGlu 262
            |||||||||   ||||||      ||||||   |||:::|||||||||||||||||||||
Db      790 GCCGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAA 846

Qy      263 LysGluProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSer 282
            |||||||||:::||||||   |||||||||:::|||||||||   |||||||||   |||
Db      847 AAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCA 906

Qy      283 SerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLys 302
            :::||||||||||||||||||   ||||||||||||||||||||||||||||||||:::
Db      907 ACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGG 966

Qy      303 LysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGlnGluAlaGluLys 322
            |||   |||||||||::::::|||   :::||||||:::||||||||||||   ||||||
Db      967 AAGCCGAAGGAGGATGAAGTGGAAGCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAA 1026

Qy      323 ValAspGluGluGluLysGluLysThrGluProAlaSerGluGlu--------------- 337
            ||||||   ||||||      |||   |||   |||||||||:::
Db     1027 GTAGACACAGAAGAAGACGGAAAGGCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAG 1086

Qy      338 ------------GlnGluProAlaGluAspThrAspGlnAlaArgLeuSerAlaAspTyr 353
                        ||||||||||||||||         :::   ||||||||||||:::|||
Db     1087 CAAGCCCACCCACAGGAGCCGGCAGAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATAT 1146

Qy      354 GluLysValGluLeuProLeuGluAspGlnValGlyAspLeuGluAlaSerSerGluGlu 373
            ||||||||||||||||||   |||:::||||||         :::      |||||||||
Db     1147 GAGAAAGTTGAGCTGCCCTCAGAGGAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAG 1206

Qy      374 LysCysAlaProLeuAlaThrGluValPheAspGluLys 386
            |||   |||||||||||||||||||||||||||||||||
Db     1207 AAACCTGCTCCGTTGGCGACAGAAGTGTTTGATGAAAAA 1245

RESULT 2
BG974884
LOCUS       BG974884              776 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602842771F1 NCI_CGAP_Mam4 Mus musculus cDNA clone IMAGE:4978744 5',
            mRNA sequence.
ACCESSION   BG974884
VERSION     BG974884.1  GI:14362521
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 776)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Priscilla Furth
```

```
          Ph.D.
           cDNA Library Preparation: Life Technologies, Inc.
           cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
           DNA Sequencing by: Incyte Genomics, Inc.
           Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM10975  row: m  column: 17
          High quality sequence stop: 753.
FEATURES             Location/Qualifiers
     source          1..776
                     /organism="Mus musculus"
                     /strain="NMRI"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4978744"
                     /clone_lib="NCI_CGAP_Mam4"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="5 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Priscilla Furth,
                     NIH Reference for transgenic model: Li et al., Cell Growth
                     and  Differentiation 7, 3-11 (1996)."
BASE COUNT      278 a    181 c    216 g    101 t
ORIGIN

Alignment Scores:
Pred. No.:              4.94e-75       Length:         776
Score:                  1041.50        Matches:        220
Percent Similarity:     89.27%         Conservative:   13
Best Local Similarity:  84.29%         Mismatches:     22
Query Match:            12.90%         Indels:         6
DB:                     13             Gaps:           3

US-09-902-432-4 (1-1596) x BG974884 (1-776)

Qy       99 GluLysAspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLys 118
            ||||||||||||   :::||||||||||||:::||||||   |||||||||||||||||||
Db        1 GAAAAAGACCGAGCTAAAGAAATGGCAGCCAGTTCCACAGTGGTTGAAGATATCACAAAG 60

Qy      119 AspGlyGlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnVal 138
            |||   |||||||||||||||||||||||||||||||||||||||||||:::||||||
Db       61 GACGAGCAGGAGGAAACACCGGAAATAATCGAACAGATCCCTGCTTCAGAGAGCAATGTG 120

Qy      139 GluGluMetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPhe 158
            |||||||||   |||   ||||||||||||||||||||||||||||||||||||||||||
Db      121 GAAGAAATGGCGCAGGCTGCTGAGTCCCAAGCTAATGACGTCGGCTTCAAGAAGGTATTT 180

Qy      159 LysPheValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrVal 178
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 AAATTTGTTGGTTTTAAATTCACGGTGAAGAAGGATAAAAACGAAAAGTCAGATACCGTC 240

Qy      179 GlnLeuLeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGly 198
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 CAGCTACTCACTGTCAAGAAGGATGAAGGCGAAGGGGCAGAAGCCTCCGTCGGAGCAGGA 300

Qy      199 AspHisGlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGlu 218
            |||||||||||||||   |||||||||   ||||||||||||||||||||||||||||||
Db      301 GACCACCAAGAGCCCGGAGTGGAGACC---GTCGGCGAATCAGCATCCAAAGAAAGTGAG 357

Qy      219 LeuLysGlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGlu 238
            |||||||||||||||||||||||||||||||||||||||   |||||||||||||||
Db      358 CTGAAGCAATCCACAGAGAAGCAAGAAGGCACCCTGAAGCAAGCACAGAGCAGCACAGAA 417

Qy      239 IleProLeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGlu 258
            ||||||||||||||||||   |||      |||||||||||      :::||||||
Db      418 ATTCCCCTTCAAGCCGAATCTGGTCAAGGGACCGAGGAAGAAGCAGCCAAAGATGGAGAA 477

Qy      259 GluLysGlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSer 278
            |||   :::||||||||||||||||||      |||||||||:::|||||||||::::::
Db      478 GAAAACCGAGAGAAAGAACCTACCAAGCCCCTAGAATCTCCGACCAGCCCTGTCAGCAAT 537

Qy      279 GluThrThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLys 298
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      538 GAGACAACATCTTCCTTCAAGAAATTCTTCACTCACGGCTGGGCCGGCTGGCGCAAGAAG 597

Qy      299 ThrSerPheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGln 318
            ||||||||||||||   |||||||||||||||||||||:::|||||||||:::||||||
Db      598 ACCAGCTTCAAGAAACCAAAGGAAGATGATCTGGAAACTTCCGAGAAGAGACAGGAGCAA 657

Qy      319 GluAlaGluLysValAspGluGluGluLys---GluLysThrGluProAla---SerGlu 336
            ||||||||||||||||||         |||      |||||||||||||||    :::|||
Db      658 GAGGCTGAAAAAGTAGACCGAGGAAGCAAGGGGACCAAGACAGAGCCAGCCCCAGCCGAG 717

Qy      337 GluGlnGlu-ProAlaGlu-AspThrAspGlnAlaArgLeuSer-AlaAspTyrGlu 354
            ||||||||| |||||||||   |||||||||||||||::::::  ||||||||||||
Db      718 GAGCAGGAAGCCTGCAGAAGGGCACAGACCAGGCCAGGATGACCAGCCGACTATGAG 774

RESULT 3
BB546856
LOCUS       BB546856              668 bp    mRNA    linear   EST 26-OCT-2001
DEFINITION  BB546856 RIKEN full-length enriched, 0 day neonate eyeball Mus
            musculus cDNA clone E130314I19 3' similar to AB020886 Mus musculus
            mRNA for SSeCKS, mRNA sequence.
ACCESSION   BB546856
VERSION     BB546856.2  GI:16447501
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 668)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
  JOURNAL   Unpublished (2001)
COMMENT     On Jul 31, 2000 this sequence version replaced gi:9618284.
            Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
             wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
             RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res. .
            10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp/) for
```

```
            further details.
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
FEATURES             Location/Qualifiers
     source          1..668
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="E130314I19"
                     /clone_lib="RIKEN full-length enriched, 0 day neonate
                     eyeball"
                     /tissue_type="eyeball"
                     /dev_stage="0 day neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAGAGCGGCCGCAACTCGAGTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. Second strand cDNA was prepared with the
                     primer adapter of sequence [5'
                     GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3']. cDNA
                     was cleaved with BamHI and XhoI. Vector: a modified
                     pBluescript KS(+) after bulk excision from Lambda FLC I."
BASE COUNT      199 a    146 c    229 g     93 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              1.01e-71        Length:         668
Score:                  1000.00         Matches:        198
Percent Similarity:     91.48%          Conservative:   6
Best Local Similarity:  88.79%          Mismatches:     17
Query Match:            12.39%          Indels:         2
DB:                     10              Gaps:           2

US-09-902-432-4 (1-1596) x BB546856 (1-668)

Qy     385 GluLysMetGluAlaHisGlnGluValValAlaGluValHisValSerThrValGluLys 404
           |||||  |||   |||||||||||||||||||||||||||||||||||||||||||||||
Db       3 GAGAAGACGGAGNCCCACCAAGAAGTTGTTGCAGAGGTCCACGTGAGCACCGTGGAGAAG 62

Qy     405 ThrGluGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValValValGluGlyThr 424
                 :::  |||  |||||   |||   |||||   |||||||||||||||:::
Db      63 ATGACGAAAGGGCAA---GGAGGAGCAGAGGTGGAAGGGGATGTGGTGGTGGAAGGATCG 119

Qy     425 GlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAlaGlu 444
           ||||||||||||||||||||||||||||||||   ||||||||||||||||||||||||
Db     120 GGAGAATCCTTGCCCCCTGAGAAACTGGCTGAGACCCAGGAGGTCCCCCAGGAAGCTGAG 179

Qy     445 ProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThrGln 464
           |||   |||||||||||||:::::|||:::||||||||||||||||||||||||||||||
Db     180 CCTGTGGAGGAGCTGATGAAGACCAAAGAAGTATGCGTCTCTGGGGGTGACCATACTCAG 239

Qy     465 LeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleValSer 484
           ||||||||||||||||||||||||||||||   |||||||||||||||||||||||||||
Db     240 CTGACAGATCTAAGTCCTGAAGAGAAGATGCTACCCAAACACCCCGAAGGCATTGTCAGT 299

Qy     485 GluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLysLys 504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     300 GAGGTGGAGATGCTGTCCTCTCAGGAGAGAATCAAGGTACAGGGAAGTCCCCTGAAGAAG 359

Qy     505 LeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArg--- 523
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     360 CTCTTCAGCAGTTCGGGCTTAAAGAAGCTCTCCGGGAAGAAGCAGAAGGGGAAGAGAGGA 419

Qy     524 GlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGlu 543
           ||||||||||||||||||||||||||||||||||||||||||    ||||||||||||||
Db     420 GGAGGCGGGGGAGATGAAGAGCCAGGAGAATACCAACACATTCAAACCGAGTCCCCAGAG 479

Qy     544 SerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThr 563
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     480 AGTGCTGACGAGCAGAAGGGAGAGAGCTCTGCCTCTTCCCCTGAAGAGCCCGAGGAGATC 539

Qy     564 ThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThr 583
              ||||||||||||||||   |||||||||||:::   ||||||||||||||
Db     540 GCGTGTCTGGAGAAGGGGCCATCGGAAGCACCCCAGGAAGCGGAAGCTGAGGAAGGAGCG 599

Qy     584 ThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMet 603
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     600 ACTTCCGACGGAGAGAAGAAAAGGGAAGGGATCACCCCCTGGGCATCCTTCAAAAAGATG 659

Qy     604 ValThrPro 606
           |||||||||
Db     660 GTGACACCC 668

RESULT 4
BI652823
LOCUS       BI652823                 661 bp    mRNA    linear   EST 12-SEP-2001
DEFINITION  603298753F1 NCI_CGAP_Mam3 Mus musculus cDNA clone IMAGE:5339244 5',
            mRNA sequence.
ACCESSION   BI652823
VERSION     BI652823.1  GI:15567059
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 661)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Lothar Hennighausen Ph.D., Chu-Xia Deng Ph.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11861  row: j  column: 13
            High quality sequence stop: 661.
FEATURES             Location/Qualifiers
     source          1..661
                     /organism="Mus musculus"
                     /strain="129,C57BL/6J,FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5339244"
                     /clone_lib="NCI_CGAP_Mam3"
                     /tissue_type="tumor, gross tissue"
                     /dev_stage="10 months"
                     /lab_host="DH10B"
                     /note="Organ: mammary; Vector: pCMV-SPORT6; Site_1: SalI;
                     Site_2: NotI; Cloned unidirectionally.  Primer: Oligo dT.
                     Library constructed by Life Technologies. Investigators
                     providing samples: Lothar Hennighausen/Chu-Xia Deng, NIH
                     Reference for transgenic model: Xu et al., Nature Genetics
                     22, 37-43 (1999)."
BASE COUNT      183 a    179 c    201 g     98 t
ORIGIN

Alignment Scores:
Pred. No.:              3.22e-64        Length:         661
Score:                  908.00          Matches:        187
Percent Similarity:     90.91%          Conservative:   13
Best Local Similarity:  85.00%          Mismatches:     20
Query Match:            11.25%          Indels:         1
DB:                     13              Gaps:           0
```

```
US-09-902-432-4 (1-1596) x BI652823 (1-661)

Qy     876 ValIleAspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSer 895
           |||||||||||||||||    |||||||||   |||||||||||||||||||||||||
Db       3 GTCATTGATGGACCAGGCA-GTCACCAGTGCCGAAGAGCGGTCCCCTTCGTGGATATCT 61

Qy     896 AlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGluGlu 915
           |||||:::|||||||||||||||||       |||   |||   |||||||||   |||
Db      62 GCTTCCATGACAGAACCTCTTGAGCACGCAGAGGGAGTGGCCACACCGCCTGTTGGAGAG 121

Qy     916 ValThrGluLysAspIleIleAlaGluGluThrProValLeuThrGlnThrLeuProGlu 935
           |||||||||||||||||    ||||||   ||||||   |||   |||||||||||
Db     122 GTCACTGAAAAAGACATCACTGCAGAAGCAACTCCTGCACTCGCCCAGACTTTACCAGGG 181

Qy     936 GlyLysAspAlaHisAspAspMetValThrSerGluValAspPheThrSerGluAlaVal 955
           ||||||||||||||||||||:::||||||||||||||||||||||||||||||||||||
Db     182 GGCAAAGATGCCCATGACGACATAGTCACCAGTGAAGTGGATTTTACCTCAGAAGCAGTG 241

Qy     956 ThrAlaThrGluThrSerGluAlaLeuArgThrGluGluValThrGluAlaSerGlyAla 975
           ||||||   ||||||:::||||||||||||   ||||||:::|||||||||||||||||
Db     242 ACAGCCGCAGAAACCACAGAGGCGCTCCGCGCTGAAGAACTTACCGAAGCATCAGGGGCA 301

Qy     976 GluGluThrThrAspMetValSerAlaValSerGlnLeuThrAspSerProAspThrThr 995
           ||||||||||||||||||||||||||||||||||||:::||||||||||||||||||||
Db     302 GAAGAGACCACAGACATGGTGTCTGCAGTTTCCCAGCTGTCCGACTCCCCGGACACCACA 361

Qy     996 GluGluAlaThrProValGlnGluValGluSerGlyValLeuAspThrGluGluGluGlu 1015
           |||||||||||||||||||||||||||||   |||:::||||||||||||||||:::|||
Db     362 GAGGAAGCCACCCCAGTTCAGGAGGTAGAGGGTGGCATGCTAGATACGGAAGAACAGGAG 421

Qy    1016 ArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysValLysGluGluSerGlnVal 1035
           ||||||||||||||:::|||||||||||||||||||||||||||||||:::|||||||||
Db     422 CGCCAGACGCAGGCCGTCCTCCAAGCCGTTGCAGACAAAGTGAAAGAGGACTCCCAGGTG 481

Qy    1036 ProAlaThrGlnThrValGlnArgThrGlySerLysAlaLeuGluLysValGluGluVal 1055
           ||||||||||||||:::||||||   |||   ||||||||||||||||||||||||||
Db     482 CCTGCAACCCAGACTCTGCAGAGAGCAGGGCCGAAAGCACTGGAGAAGGTGGAGGAGGTA 541

Qy    1056 GluGluAspSerGluValLeuAlaSerGluLysGluLysAspValMetProLysGlyPro 1075
           ||||||||||||||||||||||:::||||||||||||||||||:::|||:::||||||
Db     542 GAGGAGGACTCCGAGGTGCTGGCTACCGAGAAAGAGAAGGATGTTGTGCCGGAAGGACCC 601

Qy    1076 ValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGluThrGlyGlnAlaThrPro 1095
           ||||||||||||         ||||||||||||||||||||||||   ||||||||||||
Db     602 GTGCAGGAAGCTGAAACTGAGCATCTTGCACAGGGCTCCGAGACTGTACAGGCTACCCCA 661
```

RESULT 5
AL045541

```
LOCUS       AL045541                 674 bp    mRNA    linear   EST 29-FEB-2000
DEFINITION  DKFZp434K035_r1 434 (synonym: htes3) Homo sapiens cDNA clone
            DKFZp434K035 5', mRNA sequence.
ACCESSION   AL045541
VERSION     AL045541.1  GI:5433672
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 674)
  AUTHORS   Duesterhoeft,A., Lauber,J., Mewes,H.W., Gassenhuber,J. and Wiemann
            ,S.
  TITLE     EST (Duesterhoeft, et al.)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Duesterhoeft A
            MIPS
            Am Klopferspitz 18a D-82152 Martinsried, Germany
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz- heidelberg.de;
            sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
            consortium of the German Genome Project.
            No s1 sequence available.
            This clone (DKFZp434K035) is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin- Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES             Location/Qualifiers
     source          1..674
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="DKFZp434K035"
                     /clone_lib="434 (synonym: htes3)"
                     /tissue_type="testis"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT      213 a    139 c    212 g    109 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:               5.28e-60      Length:         674
Score:                   856.50        Matches:        170
Percent Similarity:      86.04%        Conservative:   21
Best Local Similarity:   76.58%        Mismatches:     30
Query Match:             10.61%        Indels:         1
DB:                      9             Gaps:           1

US-09-902-432-4 (1-1596) x AL045541 (1-674)

Qy     424 ThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValProGlnGluAla 443
           |||     |||:::|||   |||:::|||   |||            ||||||||||
Db       4 ACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCTCAGGAAGCC 63

Qy     444 GluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGlyAspHisThr 463
           |||||||||:::||||||:::|||      :::|||   |||||||||||   |||   |||
Db      64 GAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAGGACCCTACA 123

Qy     464 GlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGluGlyIleVal 483
           |||      ||||||||||||:::||||||   |||   |||   |||||||||:::|||
Db     124 CAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAAGGCGTTGTG 183

Qy     484 SerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySerProLeuLys 503
           ||||||||||||||||||||||||||||||:::|||||||||||||||||||||||||
Db     184 AGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGTCCACTAAAG 243

Qy     504 LysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArg 523
           ||||||||:::|||:::||||||||||||||||||||||||||||||||||||||||
Db     244 AAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA 303

Qy     524 GlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGlu 543
              |||||||||||||||||   ||||||:::         :::      :::||||||:::
Db     304 ---GGAGGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCGGAC 360

Qy     544 SerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThr 563
           |||   :::||||||||||||||||||||||||||||||||||||||||||||||||
Db     361 AGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAGATC 420

Qy     564 ThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThr 583
           ||||||||||||||||||      |||      ||||||||||||||||||||||||
Db     421 ACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCT 480

Qy     584 ThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMet 603
           ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||
Db     481 ACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAGATG 540

Qy     604 ValThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeuGlu 623
           ||||||||||||||||||||||||||||||||||||||||||||:::||||||:::
Db     541 GTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTGGAC 600

Qy     624 LysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGlu 643
           |||||||||||||||||||||||||||:::||||||   ||||||||||||:::|||
Db     601 AAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAAGAA 660

Qy     644 ValLys 645
           :::|||
Db     661 ATGAAA 666
```

```
RESULT 6
BM804332
LOCUS       BM804332                1117 bp    mRNA    linear   EST 05-MAR-2002
DEFINITION  AGENCOURT_6462209 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:5581233
            5', mRNA sequence.
ACCESSION   BM804332
VERSION     BM804332.1  GI:19121155
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1117)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM12340  row: m  column: 10
            High quality sequence stop: 623.
FEATURES             Location/Qualifiers
     source          1..1117
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5581233"
                     /clone_lib="NIH_MGC_67"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.  Library constructed by Life
                     Technologies."
BASE COUNT      361 a    257 c    339 g    159 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:              1.42e-59       Length:         1117
Score:                  854.50         Matches:        206
Percent Similarity:     66.58%         Conservative:   37
Best Local Similarity:  56.44%         Mismatches:     77
Query Match:            10.58%         Indels:         46
DB:                     14             Gaps:           10

US-09-902-432-4 (1-1596) x BM804332 (1-1117)

Qy    291 GlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSerLysGluAspAspLeuGlu 310
Db      1 GGACGCGTGGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCGAAGGAGGATGAAGTGGAA 60

Qy    311 ThrAlaGluLysArgLysGluGlnGluAlaGluLysValAspGluGluGluLysGluLys 330
Db     61 GCTTCAGAGAAGAAAAAGGAACAAGAGCCAGAAAAAGTAGACACAGAAGAAGACGGAAAG 120

Qy    331 ThrGluProAlaSerGluGlu--------------------------GlnGluProAla 341
Db    121 GCAGAGGTTGCCTCCGAGAAACTGACCGCCTCCGAGCAAGCCCACCCACAGGAGCCGGCA 180

Qy    342 GluAspThrAspGlnAlaArgLeuSerAlaAspTyrGluLysValGluLeuProLeuGlu 361
Db    181 GAAAGTGCCCACGAGCCCCGGTTATCAGCTGAATATGAGAAAGTTGAGCTGCCCTCAGAG 240

Qy    362 AspGlnValGlyAspLeuGluAlaSerSerGluGluLysCysAlaProLeuAlaThrGlu 381
Db    241 GAGCAAGTCAGTGGCTCGCAGGGACCTTCTGAAGAGAAACCTGCTCCGTTGGCGACAGAA 300

Qy    382 ValPheAspGluLysMetGluAlaHisGln---GluValValAlaGluValHisValSer 400
Db    301 GTGTTTGATGAGAAAATAGAAGTCCACCAAGAAGAGGTTGTGGCCGAAGTCCACGTCAGC 360

Qy    401 ThrValGluLysThrGluGluGlnGlyGlyGlyGlyGluAlaGluGlyGlyValVal 420
Db    361 ACCGTGGAGGAGAGAACCGAAGAGCAG----------------------AAAACGGAG 396

Qy    421 ValGluGlyThrGlyGluSerLeuProProGluLysLeuAlaGluProGlnGluValPro 440
Db    397 GTGGAAGAAACAGCAGGGTCTGTGCCAGCTGAAGAATTGGTTGAAATGGATGCAGAACCT 456

Qy    441 GlnGluAlaGluProAlaGluGluLeuMetLysSerArgGluMetCysValSerGlyGly 460
Db    457 CAGGAAGCCGAACCTGCCAAGGAGCTGGTGAAGCTCAAAGAAACGTGTGTTTCCGGAGAG 516

Qy    461 AspHisThrGlnLeuThrAspLeuSerProGluGluLysThrLeuProLysHisProGlu 480
Db    517 GACCCTACACAGGGAGCTGACCTCAGTCCTGATGAGAAGGTGCTGTCCAAACCCCCCGAA 576

Qy    481 GlyIleValSerGluValGluMetLeuSerSerGlnGluArgIleLysValGlnGlySer 500
Db    577 GGCGTTGTGAGTGAGGTGGAAATGCTGTCATCACAGGAGAGAATGAAGGTGCAGGGAAGT 636

Qy    501 ProLeuLysLysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLys 520
Db    637 CCACTAAAGAAGCTTTTACCAGCACTGGCTTNAAAAAGCTTTCTGGGAAAGAACAGAAG 696

Qy    521 GlyLysArgGlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIle-HisThrGl 540
Db    697 GGAAAAGAG---GGAGA-GGAGACGAGGAATCCGGGGAGCACACTTCAGTTCCAGGCCGA 752

Qy    540 uSerProGluSerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGlu-P 560
Db    753 TTCTCCGGACAGCCAGGAAGAGCCAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGGC 812

Qy    560 roGluGluThrThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaG 580
Db    813 CGAAGAAGACCCCTGGTCTGGAAAAAGGGCTTATC------CCAAGG---GGCCACCAGG 863

Qy    580 luGluGlyThrThr-----------------SerAspGlyGluLysLysArgGluGlyI 594
Db    864 AATGGGGAACCTGTAAAAAGGAACCTACTTTCGCAGGGGGAAAAAAAAAGAAAAAGGGG 923

Qy    594 leThrPro---TrpAlaSerPheLysLys------MetValThrProLysLysArgVal- 610
Db    924 CCACTCCCCGGGGGGCTTCTTTTCCAAAAGGAGGGGGGAGCCCCCCAAAAAAAGCGGTGT 983

Qy    611 ---------------------------------ArgArgProSerGluSerAspLysG 619
Db    984 TAAACGGGCCCTTCCGGAAAAGGGGTATAACAAAAAAGAGAACTCGGGAACACGGGGTAC 1043

Qy    619 luGluGluLeu 622
Db   1044 AAGAACGCCTC 1054

RESULT 7
D88456
LOCUS       D88456                  1153 bp    mRNA    linear   EST 01-APR-1998
DEFINITION  D88456 Mouse ICR testis (J.Tsuchida) Mus musculus cDNA clone A-12,
            mRNA sequence.
ACCESSION   D88456
VERSION     D88456.1  GI:1669580
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 1153)
  AUTHORS   Taketo,M.M., Araki,Y., Matsunaga,A., Yokoi,A., Tsuchida,J., Nishina
            ,Y., Nozaki,M., Tanaka,H., Koga,M., Uchida,K., Matsumiya, Okuyama
            ,A., Rochelle,J.M., Nishimune,Y., Matsui,M. and Seldin,M.
  TITLE     Mapping of eight testis-specific genes to mouse chromosomes
  JOURNAL   Genomics 46 (1), 138-142 (1997)
```

```
  MEDLINE   98066773
COMMENT     Contact: Junji Tsuchida
            Department of Science for Laboratory Animal Experimentation
            Research Institute for Microbial Diseases, Osaka University
            3-1 Yamadaoka, Suita, Osaka 565, Japan
            Tel: 06-879-8338
            Fax: 06-879-8339
            Email: tsuchida@biken.osaka-u.ac.jp.
FEATURES             Location/Qualifiers
     source          1. .1153
                     /organism="Mus musculus"
                     /strain="ICR"
                     /db_xref="taxon:10090"
                     /clone="A-12"
                     /clone_lib="Mouse ICR testis (J.Tsuchida)"
                     /sex="male"
                     /tissue_type="testis"
BASE COUNT      367 a    285 c    299 g    201 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                1.78e-59      Length:         1153
Score:                    853.50        Matches:        182
Percent Similarity:       80.08%        Conservative:   19
Best Local Similarity:    72.51%        Mismatches:     43
Query Match:              10.57%        Indels:         8
DB:                       14            Gaps:           2

US-09-902-432-4 (1-1596) x D88456 (1-1153)

Qy    1351 GlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluValGlnValGlnSer 1370
           ||||||||||||||||||||||||||||||    ||||||||||||||||||||||
Db       3 GGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGATTGCATAGAGGTTCAAGTTCAAAGC 62

Qy    1371 LeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIleGluThrValVal 1390
               ||||||   ||||||||||||   ||||||||:::||||||   |||||||||
Db      63 TCAGACACACCAGTCACTCAAACAACCGAAGCTGTGAAAAAGGTCGAAGAAACTGTGGCA 122

Qy    1391 IleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeuProAlaGluLys 1410
               ||||||      ||||||   ||||||   ||||||    ||||||||||||||
Db     123 ACTTCAGAGATGGATGAAAGTTTGGAGTGTGCAGGTGCGCAATCATTACCAGCTGAGAAG 182

Qy    1411 SerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThrValProLeuGly 1430
              |||   ||||||||:::   ||||||||||||   |||||||||||||||
Db     183 CTCTCCGAAACCGGTGGCTACGGGACTCTTCAGCATGGAGAGGACACCGTGCCCCAGGNG 242

Qy    1431 ProGluSerGlnAlaGluSerIleProIleIleValThrProAlaProGluSerThrLeu 1450
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
Db     243 CCTGAGTCTCAGGCAGAGTCCATCCCAATAATAGTAACTCCTGCTCCTGAAAGCATCCTA 302

Qy    1451 HisProAspLeuGlnGlyGluIleSerAlaSerGlnArgGluArgSerGluGluGluAsp 1470
           |||   |||||||||   |||:::||||||||||||::::::||||||:::|||::::::
Db     303 CATTCTGACCTTCAAAGAGAAGTGAGCGCATCCCAGAAACAGAGATCAGATGAAGATAAC 362

Qy    1471 LysProAspAlaGlyProAspAlaAspGlyLysGluSerThrAlaIleGluLysValLeu 1490
           ||||||||||||||||||||||||   ||||||||||||   |||   ||||||:::|||
Db     363 AAGCCAGATGCTGGTCCTGATGCTGCCGGCAAGGAGAGTGCAGCAAGAGAGAAAATCCTC 422

Qy    1491 LysAlaGluProGluIleLeuGluLeuGluSerLysSerAsnLysIleValLeuAsnVal 1510
           :::||||||||||||||||||||||||||||||||||||||||||||||||||   :::|||
Db     423 AGGGCTGAACCTGAGATCTTGGAACTTGAGAGTAAGAGCAATAAGATTGTCCAGAGTGTC 482

Qy    1511 IleGlnThrAlaValAspGlnPheAlaArgThrGluThrAlaProGluThrHisAlaTyr 1530
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     483 ATCCAGACAGCCGTCGACCAGTTTGCACGTACAGAAACAGCCCCCGAAACCCACGCTTCT 542

Qy    1531 AspSerGlnThrGlnValProAlaCysArgLeuAspSerArgGluProAsnArgCysTrp 1550
           |||   |||   |||||||||          :::   :::||||||   ||||||||||||
Db     543 GATTTACAGAATCAGGTTCCTGTG------ATGCA-GCCAGGGAGCACAACAGATGCTGG 595

Qy    1551 ThrLysMetLys---------------AspAlaLysMetLysHisProValProGlnPro 1565
           ||||||||||||                  ||||||:::|||   ||||||   |||
Db     596 ACAAAGATGAAAGCGACCTTCAAGTCTCCCCCAAGATGGAACACTCAGTGCCGTAGCCC 655

Qy    1566 ArgGluAspLeuGlnValLeuThrValLeuGluAlaTrpAlaGlnProArgLysCysLeu 1585
           |||:::||||||:::   |||   ||||||:::|||   |||:::   :::|||
Db     656 AGGAAGGACTTGCGGTTTCTGATAGTTCTGAAGGCATGAGCAAGGCTTCAGAAATGATCA 715

Qy    1586 ProArgLeuGlnLeuLysAlaProValSerLys 1596
           ||||||||||||||||||   ||||||||||||
Db     716 CCACGCTTGCAGTTGAAAGTGCCAGTGTCAAAG 748

RESULT 8
AI225639
LOCUS       AI225639                610 bp    mRNA    linear   EST 29-OCT-1998
DEFINITION  uj13g09.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:1907968 5' similar to TR:Q62766 Q62766 SSECKS. ;, mRNA
            sequence.
ACCESSION   AI225639
VERSION     AI225639.1  GI:3808692
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 610)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:976164
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 460.
FEATURES             Location/Qualifiers
     source          1. .610
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1907968"
                     /clone_lib="Sugano mouse kidney mkia"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraIII
                     (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a  DraIII adaptor [TGTTGGCCTACTGG], digested
                     and cloned into distinct DraIII sites of the pME18S-FL3
                     vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb.  Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers  for
                     sequencing: 5' end primer CTTCTGCTCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGAGCACA."
BASE COUNT      172 a    163 c    185 g     89 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                3.7e-57       Length:         610
Score:                    821.00        Matches:        167
Percent Similarity:       88.61%        Conservative:   12
```

```
Best Local Similarity:  82.67%        Mismatches:    23
Query Match:            10.17%        Indels:         0
DB:                     9             Gaps:           0

US-09-902-432-4 (1-1596) x AI225639 (1-610)

Qy    923 AlaGluGluThrProValLeuThrGlnThrLeuProGluGlyLysAspAlaHisAspAsp 942
          ||||||   ||||||    |||    |||||||||||    ||||||||||||||||||||
Db      5 GCAGAAGCAACTCCTGCACTCGCCCAGACTTTACCAGGGGGCAAAGATGCCCATGACGAC 64

Qy    943 MetValThrSerGluValAspPheThrSerGluAlaValThrAlaThrGluThrSerGlu 962
          :::|||||||||||||||||||||||||||||||||||||||||||    ||||||:::|||
Db     65 ATAGTCACCAGTGAAGTGGATTTTACCTCAGAAGCAGTGACAGCCGCAGAAACCACAGAG 124

Qy    963 AlaLeuArgThrGluGluValThrGluAlaSerGlyAlaGluGluThrThrAspMetVal 982
          |||||||||   ||||||:::||||||||||||||||||||||||||||||||||||||||
Db    125 GCGCTCCGCGCTGAAGAACTTACCGAAGCATCAGGGGCAGAAGAGACCACAGACATGGTG 184

Qy    983 SerAlaValSerGlnLeuThrAspSerProAspThrThrGluGluAlaThrProValGln 1002
          ||||||||||||||||||:::||||||||||||||||||||||||||||||||||||||||
Db    185 TCTGCAGTTTCCCAGCTGTCCGACTCCCCGGACACCACAGAGGAAGCCACCCCAGTTCAG 244

Qy   1003 GluValGluSerGlyValLeuAspThrGluGluGluGluArgGlnThrGlnAlaIleLeu 1022
          |||||||||   ||||:::||||||||||||||||:::|||||||||||||||||:::|||
Db    245 GAGGTAGAGGGTGGCATGCTAGATACGGAAGAACAGGAGCGCCAGACGCAGGCCGTCCTC 304

Qy   1023 GlnAlaValAlaAspLysValLysGluGluSerGlnValProAlaThrGlnThrValGln 1042
          ||||||||||||||||||||||||||||:::||||||||||||||||||||||||:::|||
Db    305 CAAGCCGTTGCAGACAAAGTGAAAGAGGACTCCCAGGTGCCTGCAACCCAGACTCTGCAG 364

Qy   1043 ArgThrGlySerLysAlaLeuGluLysValGluGluValGluGluAspSerGluValLeu 1062
          |||    |||   |||||||||||||||||||||||||||      ||||||   |||||||||
Db    365 AGAGCAGGGCCGAAAGCACTGGAGAAGGTGGAGGAGGTAGAGAGGACTTCGAGGTGCTG 424

Qy   1063 AlaSerGluLysGluLysAspValMetProLysGlyProValGlnGluAlaGlyAlaGlu 1082
          |||:::|||||||||||||||||:::|||:::||||||||||||||||||      |||
Db    425 GCTACCGAGAAAGAGAAGGATGTTGTGCCGGAAGGACCCGTGCAGGAAGCTNGAACTGAG 484

Qy   1083 HisLeuAlaGlnGlySerGluThrGlyGlnAlaThrProGluSerLeuGluValProGlu 1102
          ||||||||||||||   |||||||||   |||||||||||||||||||    ||||||
Db    485 CATCTTGCACAGGGCTTCGAGACTGGACCAGCTACCCCAGAGAGCCTTTGAGTTCCTGGA 544

Qy   1103 ValThrAlaAspValAspHisValAlaThrCysGlnValIleLysLeuGlnGlnLeuMet 1122
          ||||||   |||||||||          |||||||||||||||||    |||||||||||
Db    545 GTCACAGAGGATGTAGACCGTGCCACCACATGCCAGGTTATCAAGCACCAGCAGCTGATG 604

Qy   1123 GluGln 1124
          ||||||
Db    605 GAACAG 610

RESULT 9
BE618075
LOCUS       BE618075               646 bp    mRNA    linear   EST 20-OCT-2000
DEFINITION  601462263F1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3865910 5',
            mRNA sequence.
ACCESSION   BE618075
VERSION     BE618075.1  GI:9889013
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 646)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM9609  row: m  column: 15
             High quality sequence stop: 646.
FEATURES             Location/Qualifiers
     source          1..646
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3865910"
                     /clone_lib="NIH_MGC_67"
                     /tissue_type="retinoblastoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.75 kb.  Library constructed by Life
                     Technologies."
BASE COUNT      212 a    128 c    207 g     99 t
ORIGIN

Alignment Scores:
Pred. No.:              6.05e-56      Length:        646
Score:                  806.50        Matches:       164
Percent Similarity:     86.92%        Conservative:  22
Best Local Similarity:  76.64%        Mismatches:    25
Query Match:            9.99%         Indels:        4
DB:                     10            Gaps:          1

US-09-902-432-4 (1-1596) x BE618075 (1-646)

Qy    504 LysLeuPheSerSerSerGlyLeuLysLysLeuSerGlyLysLysGlnLysGlyLysArg 523
          |||||||||:::|||:::||||||||||||||||||||||||||||||||||||||||||
Db      1 AAGCTTTTTACCAGCACTGGCTTAAAAAAGCTTTCTGGAAAGAAACAGAAAGGGAAAAGA 60

Qy    524 GlyGlyGlyGlyAspGluGluProGlyGluTyrGlnHisIleHisThrGluSerProGlu 543
             |||||||||||||||||    ||||||:::        :::       :::||||||:::
Db     61 ---GGAGGAGGAGACGAGGAATCAGGGGAGCACACTCAGGTTCCAGCCGATTCTCCGGAC 117

Qy    544 SerAlaAspGluGlnLysGlyGluSerSerAlaSerSerProGluGluProGluGluThr 563
          |||    :::|||||||||||||||||||||||||||||||||||||||||||||||
Db    118 AGCCAGGAGGAGCAAAAGGGCGAGAGCTCTGCCTCATCCCCTGAGGAGCCCGAGGAGATC 177

Qy    564 ThrCysLeuGluLysGlyProLeuGluAlaProGlnAspGlyGluAlaGluGluGlyThr 583
          ||||||||||||||||||       |||      |||||||||||||||||||||||||
Db    178 ACGTGTCTGGAAAAGGGCTTAGCCGAGGTGCAGCAGGATGGGGAAGCTGAAGAAGGAGCT 237

Qy    584 ThrSerAspGlyGluLysLysArgGluGlyIleThrProTrpAlaSerPheLysLysMet 603
          ||||||||||||||||||||||||||||||:::||||||||||||||||||||||||||||
Db    238 ACTTCCGATGGAGAGAAAAAAAGAGAAGGTGTCACTCCCTGGGCATCATTCAAAAAGATG 297

Qy    604 ValThrProLysLysArgValArgArgProSerGluSerAspLysGluGluGluLeu-Gl 623
          ||||||||||||||||||||||||||||||||||||||||||||||:::||||||  ::
Db    298 GTGACGCCCAAGAAGCGTGTTAGACGGCCTTCGGAAAGTGATAAAGAAGATGAGCTGGGA 357

Qy    623 uLysValLysSerAlaThrLeuSerSerThrAspSerThrValSerGluMetGlnAspGl 643
          :||||||||||||||||||||||||||||||:::||||||    ||||||||||||:::||
Db    358 CAAGGTCAAGAGCGCTACCTTGTCTTCCACCGAGAGCACAGCCTCTGAAATGCAAGAAGA 417

Qy    643 uValLysThrValGly-GluGluGlnLysProGluGluProLysArgArgValAspThrS 663
          |:::|||             ||||||   ||||||||||||||||||||:::||||||||||
Db    418 AATGAAACGGGAGCGTGGAAGAGCCAAAGCCGGAAGAACCAAAGCGCAAGGTGGATACCT 477

Qy    663 erValSerTrpGluAlaLeuIleCysValGlySerSerLysLysArgAlaArgLysAlaS 683
          ||||||||||||||||||||||||||||||||||||||||||||||||||:::     |
Db    478 CAGTATCTTGGGAAGCTTTAATTTGTGTGGGATCATCCAAGAAAAGAGCAAGGAGAGGGT 537

Qy    683 erSerSerAspAspGluGlyGlyProArgThrLeuGlyGlyAspSerHisArgAlaGluG 703
          ||||||||||:::||||||||||||:::     :::|||||||||       :::|||:::|
Db    538 CCTCTTCTGATGAGGAAGGGGGACCAAAAGCAATGGAGGAGACCACCAGAAAGCTGATG 597

Qy    703 luAlaSerLysAspLysGluAlaGlyThrAspAlaVal 715
          |||||    |||    :::|||   |||||||||     :::
Db    598 AGGCGGAAAAGA--CAAGAGACGGGGACAGACGGGATC 633
```

```
RESULT 10
AI796191/c
LOCUS       AI796191                 796 bp    mRNA    linear   EST 20-DEC-1999
DEFINITION  wh43g05.x1 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGE:2383544 3'
            similar to TR:O00498 O00498 MYASTHENIA GRAVIS AUTOANTIGEN GRAVIN ;,
            mRNA sequence.
ACCESSION   AI796191
VERSION     AI796191.1  GI:5361654
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 796)
  AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
  TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
            Emmert-Buck, M.D., Ph.D.
             cDNA Library Preparation: M. Bento Soares, Ph.D.
             cDNA Library Arrayed by: Greg Lennon, Ph.D.
             DNA Sequencing by: Washington University Genome Sequencing Center
             Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1103   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 458.
FEATURES             Location/Qualifiers
     source          1. .796
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2383544"
                     /clone_lib="NCI_CGAP_Kid11"
                     /lab_host="DH10B"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site_1: Not I; Site_2: Eco RI;
                     Plasmid DNA from the normalized library NCI_CGAP_Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502855). Subtraction by Bento Soares and M.
                     Fatima Bonaldo. "
BASE COUNT      140 a    216 c    173 g    266 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:               4.04e-54       Length:        796
Score:                   785.50         Matches:       173
Percent Similarity:      70.61%         Conservative:  24
Best Local Similarity:   62.01%         Mismatches:    54
Query Match:             9.73%          Indels:        28
DB:                      9              Gaps:          6

US-09-902-432-4 (1-1596) x AI796191 (1-796)

Qy     55 AsnGlyGlnLeuSerSerValAsnGlyValAlaGluGln--------------------- 67
             |||||   ||||||:::::  ||||||||   |||
Db    796 AATGGTCTGCTGTCNACCATCTATGGCGTAGCTGGACAGATGAGCTCAGCCTCGAGGAGG 737

Qy     68 ---------------------GlyAspValHisValGlnGluGluAsnGlnGluGlyGln 80
                                           :::     ||||||        :::
Db    736 GTAACCTAGATGTCCAGGAAGAGTCCCTGTCGGTCAAGGAGTCCTAATCAGCAT---GAA 680

Qy     81 GluGluGluValValAspGluAspValGlyGlnArgGluSerGluAspValArgGluLys 100
          ||||||||||||:::           |||      :::||||||||||||    ::::::
Db    679 GAAGAAGAAGTCATTGTCCGGCAGGTTGATCACATAGACTCTGAAGATGTGAGCGACAGA 620

Qy    101 AspArgValGluGluMetAlaAlaAsnSerThrAlaValGluAspIleThrLysAspGly 120
          |||      ||||||          |||      |||  |||||||||    ||||||
Db    619 GACTCCGATACAGAGATGACTACTAAGTCAGCGGTAGTTCACGACATCACAGATGATGGG 560

Qy    121 GlnGluGluThrSerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGlu 140
          ||||||||||||     :::|||||||||||||||||:::||||||:::|||:::||||||
Db    559 CAGGAGGAGACACCCGATATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAG 500

Qy    141 MetValGlnProAlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPhe 160
          :::   ||||||   ||||||||||||||||||:::||||||||||||||||||||||||
Db    499 CTAACACAACCCACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTT 440

Qy    161 ValGlyPheLysPheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeu 180
          ||||||||||||||||||||||||||||||||||||||||   ||||||   ||||||||||||||
Db    439 GTTGGCTTTAAATTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTA 380

Qy    181 LeuThrValLysLysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHis 200
          ||||||||||||||||||||||||||||||||||||||||          ||||||||||||||
Db    379 CTCACTGTGAAGAAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCAC 329

Qy    201 GlnGluProSerValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLys 220
          |||:::||||||:::                |||||:::||||||||||||||||||   |||
Db    328 CAGGACCCCAGCCTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAA 275

Qy    221 GlnSerThrGluLysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIlePro 240
          ||||||||||||||   |||   |||||||||:::|||||||||         ||||||
Db    274 CAATCTACAGAGAAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCT 215

Qy    241 -LeuGlnAlaGluSerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLy 260
             ||||||||||||   ||||||      ||||||   |||:::|||||||||||||
Db    214 CCCCCAAGCGGAATCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAA 158

Qy    260 sGlnGluLysGluProThrLysSerProGluSerProSerSerProValAsnSerGluTh 280
          ||||||||||||||||||:::||||||   ||||||||||:::|||   |||   ||||||||
Db    157 ACAAGAAAAAGAACCTAGCAAGTCTGCAGAATCTCCGACTAGTTCCGTGACCAGTGAAAC 98

Qy    280 rThrSerSerPheLysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSe 300
          |   |||:::||||||||||||||||||   ||||||||||||||||||||||||   ||
Db     97 AGGATCAACCTTCAAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGCCCAG 38

Qy    300 rPheLysLysSerLysGluAspAspLeuGluThrAlaGluLysArgLysGluGln 318
          ||||:::|||   ||||||                  |||:::||||||   |||
Db     37 TTTCAGGAAGCCGAAGGAG------------------GAAAGGAGGAAGCTTCAG 1

RESULT 11
BQ717711
LOCUS       BQ717711                 873 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8234483 Lupski_sympathetic_trunk Homo sapiens cDNA clone
            IMAGE:6188718 5', mRNA sequence.
ACCESSION   BQ717711
VERSION     BQ717711.1  GI:21856608
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 873)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James R. Lupski
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13584  row: m  column: 07
            High quality sequence stop: 627.
FEATURES             Location/Qualifiers
     source          1. .873
```

```
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:6188718"
                        /clone_lib="Lupski_sympathetic_trunk"
                        /sex="male"
                        /tissue_type="sympathetic trunk"
                        /dev_stage="adult, 16 yr"
                        /lab_host="DH10B"
                        /note="Vector: pCMV-SPORT6 (Life Technologies); Site_1:
                        NotI; Site_2: SalI; cDNA made by oligo-dT priming.
                        Directionally cloned using the following adaptors:
                        5'-TCGACCCACGCGTCCG-3' and
                        5'-GACTAGTTCTAGATCGCGAGCGGCCGCCCT(15)-3'. Size selected >
                        1 kb for average insert length 1.9 kb. This is a primary
                        library, non-amplified. Library constructed by Life
                        Technologies and donated  by J. Lupski, M.D./Ph.D. (Baylor
                        College of Medicine); available through Life
                        Technologies."
BASE COUNT      269 a    202 c    262 g    139 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:               8.72e-54      Length:         873
Score:                   782.00        Matches:        179
Percent Similarity:      74.55%        Conservative:   26
Best Local Similarity:   65.09%        Mismatches:     56
Query Match:             9.69%         Indels:         14
DB:                      14            Gaps:           5

US-09-902-432-4 (1-1596) x BQ717711 (1-873)

Qy     740 GlyGluGlyValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSer 759
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1 GGGGAGGGCGTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCA 60

Qy     760 LysSerLysLeuGluGluLysAlaGluAsp------------SerSerValGluGlnLeu 775
           ||||||||||||||||||||:::||||||             |||    ||||||
Db      61 AAGTCCAAGCTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCC 120

Qy     776 SerThrGluIleGluProSerArgGluGluSerTrpValSerIleLysLysPheIlePro 795
           :::   :::    |||||   :::|||||||||||||||||||||||||||||||||||
Db     121 ACTCCAGACACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCT 180

Qy     796 GlyArgArgLysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGly 815
           ||||||||||||||||||   ||||||||||||||||||||    |||||||||:::|||
Db     181 GGACGAAGGAAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGG 240

Qy     816 ProValGluIleAsnGluAspAspProAsnValProAlaValValProLeuSerGluTyr 835
           |||         ||||||||||||   :::|||||||||||||||||||||||||||||||
Db     241 CCAACAGGGGCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTAT 300

Qy     836 AsnAlaValGluArgGluLysMetGlu---------AlaGlnGlyAsnThrGluLeuPro 852
           :::||||||||||||||||||||||||         ||||||   :::   |||   |||
Db     301 GATGCTGTAGAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCC 360

Qy     853 GlnLeuLeuGlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrVal 872
           :::         |||      ||||||:::|||||||||::::::   ||||||    :::
Db     361 GAGCAGAAGGCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATG 420

Qy     873 SerValAlaValIleAspGlyThrArgAlaValThrSerValGluGluArgSerProSer 892
           :::   ||||||   |||||||||||||||   |||   :::|||||||||||||||||||
Db     421 GCAGCAGCTGTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCT 480

Qy     893 TrpIleSerAlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProPro 912
           ||||||||||||||||||||||||||||||||||||||||||         ||||||
Db     481 TGGATATCTGCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTA 540

Qy     913 ValGluGluValThrGluLysAspIleIleAla---GluGluThrProValLeuThrGln 931
              |||||||||   |||::::::::||||||   ||||||   |||   :::|||:::
Db     541 ACTGAGGAGGTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCCCCCACGGTTACTGAA 600

Qy     932 ThrLeuProGluGlyLysAspAlaHisAspAspMetValThrSerGluValAspPheThr 951
              |||||||||   ::::::|||      |||   |||   ||||||   :::   |||
Db     601 CCTCTGCCAGAGAACAGAGAGGCCCGGGGCGACACGGTCGTTAGTGAGGCGGAATTGACC 660

Qy     952 SerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThr---GluGlu-ValTh 970
              |||||||||||||||   |||||||:::   |||:::          ||||||   ||
Db     661 CCCGAAGCTGTGACAGCTGCAGAAACTGCAGGGGCCATTGGGGTGCCCGAAGAAAGGAAC 720

Qy     970 rGluAlaSerGly-AlaGluGluThrThr---AspMetValSerAlaValSerGlnLeu- 988
           ||||||||||||| ||||||||||||||      :::            ||||||||||||
Db     721 CGAAGCATCTGGCTGCTGAAGAGACAACAGNAAATGGGTGTCAGCAAGTCTCCCAGTTAA 780

Qy     989 ThrAspSerProAspThr-ThrGluGluAlaThrPro 1000
           |||||||||||||||||| ||||||||||   ||||||
Db     781 ACCGACTCCCCAGACACCCACAGAGGAGGGCACTCCC 817

RESULT 12
BQ221404
LOCUS       BQ221404                 924 bp    mRNA    linear   EST 02-MAY-2002
DEFINITION  AGENCOURT_7561202 NIH_MGC_72 Homo sapiens cDNA clone IMAGE:6054858
            5', mRNA sequence.
ACCESSION   BQ221404
VERSION     BQ221404.1  GI:20402804
KEYWORDS    EST.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 924)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC/DCTD/DTP
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM13314  row: c  column: 19
            High quality sequence stop: 618.
FEATURES             Location/Qualifiers
     source          1..924
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6054858"
                     /clone_lib="NIH_MGC_72"
                     /tissue_type="melanotic melanoma"
                     /lab_host="DH10B (phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 2 kb.   Library constructed by Life
                     Technologies."
BASE COUNT      286 a    221 c    270 g    147 t
ORIGIN

Alignment Scores:
Pred. No.:               2.63e-53      Length:         924
Score:                   776.50        Matches:        159
Percent Similarity:      77.59%        Conservative:   21
Best Local Similarity:   68.53%        Mismatches:     45
Query Match:             9.62%         Indels:         7
DB:                      14            Gaps:           2

US-09-902-432-4 (1-1596) x BQ221404 (1-924)

Qy     703 GluAlaSerLysAspLysGluAlaGlyThrAspAlaValProAlaSerThrGlnGluGln 722
              ||||||   ||||||||||||   ||||||||||   :::   |||   :::||||||
Db       3 GAGGCCGGAAAAGACAAAGAGACGGGGACAGACGGGATCCTTGCTGGTTCCCAAGAACAT 62

Qy     723 AspGlnAlaGlnGlySerSerSerProGluProAlaGlySerProSerGluGlyGluGly 742
              |||   ||||||||||||||||||||||   ||||||||||||:::||||||||||||
Db      63 GATCCAGGGCAGGGAAGTTCCTCCCCGGAGCAAGCTGGAAGCCCTACCGAAGGGGAGGGC 122
```

```
Qy     743 ValSerThrTrpGluSerPheLysArgLeuValThrProArgLysLysSerLysSerLys 762
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     123 GTTTCCACCTGGGAGTCATTTAAAAGGTTAGTCACGCCAAGAAAAAAATCAAAGTCCAAG 182

Qy     763 LeuGluGluLysAlaGluAsp------------SerSerValGluGlnLeuSerThrGlu 778
           ||||||||||||:::||||||            |||  ||||||      :::   :::
Db     183 CTGGAAGAGAAAAGCGAAGACTCCATAGCTGGGTCTGGTGTAGAACATTCCACTCCAGAC 242

Qy     779 IleGluProSerArgGluGluSerTrpValSerIleLysLysPheIleProGlyArgArg 798
              ||||||    :::||||||||||||||||||||||||||||||||||||||||||||
Db     243 ACTGAACCCGGTAAAGAAGAATCCTGGGTCTCAATCAAGAAGTTTATTCCTGGACGAAGG 302

Qy     799 LysLysArgAlaAspGlyLysGlnGluGlnAlaThrValGluAspSerGlyProValGlu 818
           |||||||||   ||||||||||||||||||||||    |||||||||:::||||||
Db     303 AAGAAAAGGCCAGATGGGAAACAAGAACAAGCCCCTGTTGAAGACGCAGGGCCAACAGGG 362

Qy     819 IleAsnGluAspAspProAsnValProAlaValValProLeuSerGluTyrAsnAlaVal 838
              |||||||||||   :::||||||||||||||||||||||||||||||:::||||||
Db     363 GCCAACGAAGATGACTCTGATGTCCCGGCCGTGGTCCCTCTGTCTGAGTATGATGCTGTA 422

Qy     839 GluArgGluLysMetGlu---------AlaGlnGlyAsnThrGluLeuProGlnLeuLeu 855
           ||||||||||||||||||         ||||||   :::   |||   |||:::
Db     423 GAAAGGGAGAAAATGGAGGCACAGCAAGCCCAAAAAAGCGCAGAGCAGCCCGAGCAGAAG 482

Qy     856 GlyAlaValTyrValSerGluGluLeuSerLysThrLeuValHisThrValSerValAla 875
             |||     ||||||:::|||||||||::::::   ||||||   ::::::   |||
Db     483 GCAGCCACTGAGGTGTCCAAGGAGCTCAGCGAGAGTCAGGTTCATATGATGGCAGCAGCT 542

Qy     876 ValIleAspGlyThrArgAlaValThrSerValGluGluArgSerProSerTrpIleSer 895
           |||    |||||||||||||   |||   :::||||||||||||||||||||||||||||
Db     543 GTCGCTGACGGGACGAGGGCAGCTACCATTATTGAAGAAAGGTCTCCTTCTTGGATATCT 602

Qy     896 AlaSerValThrGluProLeuGluHisThrAlaGlyGluAlaMetProProValGluGlu 915
           ||||||||||||||||||||||||                ||||||           ||||||
Db     603 GCTTCAGTGACAGAACCTCTTGAACAAGTAGAAGCTGAAGCCGCACTGTTAACTGAGGAG 662

Qy     916 ValThrGluLysAspIleIleAlaGluGluThrPro 927
           |||   |||:::::::::||||||||||||   |||
Db     663 GTATTGGAAAGAGAAGTAATTGCAGAAGAAGAACCC 698
```

RESULT 13
AI156819

```
LOCUS       AI156819               530 bp    mRNA    linear   EST 30-SEP-1998
DEFINITION  ui44c01.y1 Sugano mouse embryo mewa Mus musculus cDNA clone
            IMAGE:1885248 5' similar to TR:Q62766 Q62766 SSECKS. ;, mRNA
            sequence.
ACCESSION   AI156819
VERSION     AI156819.1  GI:3685288
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 530)
  AUTHORS   Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
            Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
            Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
            Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
            Waterston,R.
  TITLE     The WashU-HHMI Mouse EST Project
  JOURNAL   Unpublished (1996)
COMMENT     Contact: Marra M/Mouse EST Project
            WashU-HHMI Mouse EST Project
            Washington University School of MedicineP
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:969572
            Possible reversed clone: similarity on wrong strand
            Seq primer: custom primer used
            High quality sequence stop: 447.
FEATURES             Location/Qualifiers
     source          1..530
                     /organism="Mus musculus"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:1885248"
                     /clone_lib="Sugano mouse embryo mewa"
                     /dev_stage="embryo, 14 dpc"
                     /lab_host="DH10B"
                     /note="Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG);
                     Site_2: DraIII (CACCATGTG); 1st strand cDNA was primed
                     with an oligo(dT) primer  [ATGTGGCCTTTTTTTTTTTTTTTTT];
                     double-stranded cDNA was ligated to a  DraIII adaptor
                     [TGTTGGCCTACTGG], digested and cloned into distinct DraIII
                     sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site
                     CACCATGTG). XhoI should be used to isolate the cDNA
                     insert.  Size selection was performed to exclude fragments
                     <1.5kb.  Library constructed by Dr. Sumio Sugano
                     (University of Tokyo Institute of Medical Science).
                     Custom primers  for sequencing: 5' end primer
                     CTTCTGCTCTAAAAGCTGCG and 3' end primer
                     CGACCTGCAGCTCGAGCACA."
BASE COUNT      151 a     141 c     164 g     74 t
ORIGIN

Alignment Scores:
Pred. No.:              2.84e-51        Length:         530
Score:                  748.00          Matches:        152
Percent Similarity:     92.61%          Conservative:   11
Best Local Similarity:  86.36%          Mismatches:     13
Query Match:            9.27%           Indels:         0
DB:                     9               Gaps:           0

US-09-902-432-4 (1-1596) x AI156819 (1-530)

Qy     950 PheThrSerGluAlaValThrAlaThrGluThrSerGluAlaLeuArgThrGluGluVal 969
           ||||||||||||||||||||||||   ||||||:::||||||||||||   ||||||:::
Db       3 TTTACCTCAGAAGCAGTGACAGCCGCAGAAACCACAGAGGCGCTCCGCGCTGAAGAACTC 62

Qy     970 ThrGluAlaSerGlyAlaGluGluThrThrAspMetValSerAlaValSerGlnLeuThr 989
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||:::
Db      63 ACCGAAGCATCAGGGGCAGAAGAGACCACAGACATGGTGTCTGCAGTTTCCCAGCTGTCC 122

Qy     990 AspSerProAspThrThrGluGluAlaThrProValGlnGluValGluSerGlyValLeu 1009
           |||||||||||||||||||||||||||||||||||||||||||||||||   |||:::|||
Db     123 GACTCCCCGGACACCACAGAGGAAGCCACCCCAGTTCAGGAGGTAGAGGGTGGCATGCTA 182

Qy    1010 AspThrGluGluGluGluArgGlnThrGlnAlaIleLeuGlnAlaValAlaAspLysVal 1029
           ||||||||||||:::|||||||||||||||||||:::||||||||||||||||||||||||
Db     183 GATACGGAAGAACAGGAGCGCCAGACGCAGGCCGTCCTCCAAGCCGTTGCAGACAAAGTG 242

Qy    1030 LysGluGluSerGlnValProAlaThrGlnThrValGlnArgThrGlySerLysAlaLeu 1049
           ||||||:::||||||||||||||||||||||||:::||||||   |||   |||||||||
Db     243 AAAGAGGACTCCCAGGTGCCTGCAACCCAGACTCTGCAGAGAGCAGGGCCGAAAGCACTG 302

Qy    1050 GluLysValGluGluValGluGluAspSerGluValLeuAlaSerGluLysGluLysAsp 1069
           ||||||||||||||||||||||||||||||||||||||||||:::|||||||||||||||
Db     303 GAGAAGGTGGAGGAGGTAGAGGAGGACTCCGAGGTGCTGGCTACCGAGAAAGAGAAGGAT 362

Qy    1070 ValMetProLysGlyProValGlnGluAlaGlyAlaGluHisLeuAlaGlnGlySerGlu 1089
           |||:::|||:::||||||||||||||||||      ||||||||||||||||||||||||
Db     363 GTTGTGCCGGAAGGACCCGTGCAGGAAGCTGAAACTGAGCATCTTGCACAGGGCTCCGAG 422

Qy    1090 ThrGlyGlnAlaThrProGluSerLeuGluValProGluValThrAlaAspValAspHis 1109
           |||   |||||||||||||||||||||||||||||||||||||||   |||||||||
Db     423 ACTGTACAGGCTACCCCAGAGAGCCTTGAAGTTCCTGAAGTCACAGAGGATGTAGACCGT 482

Qy    1110 ValAlaThrCysGlnValIleLysLeuGlnGlnLeuMetGluGlnAla 1125
                 ||||||||||||||||||   ||||||||||||||||||
Db     483 GCCACCACATGCCAGGTTATCAAGCACCAGCAGCTGATGGAACAGGCT 530
```

```
RESULT 14
BQ443215
LOCUS       BQ443215               679 bp    mRNA    linear   EST 29-MAY-2002
DEFINITION  UI-M-EV0-bxg-o-07-0-UI.r1 NIH_BMAP_EV0 Mus musculus cDNA clone
            IMAGE:5707998 5', mRNA sequence.
ACCESSION   BQ443215
VERSION     BQ443215.1  GI:21246327
KEYWORDS    EST.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 679)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. James Lin, Univeristy of Iowa
             cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
             cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
             DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
             Clone Distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
             This clone was contributed by the Brain Molecular Anatomy Project
            (BMAP)
            Seq primer: pYX-5.
FEATURES             Location/Qualifiers
     source          1..679
                     /organism="Mus musculus"
                     /strain="C57BL/6"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5707998"
                     /clone_lib="NIH_BMAP_EV0"
                     /tissue_type="whole brain"
                     /dev_stage="embryo 15.5 dpc"
                     /lab_host="DH10B (T1 phage resistant)"
                     /note="Organ: brain; Vector: pYX-Asc; Site_1: EcoR I;
                     Site_2: Not I; The library was constructed according to
                     Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
                     1996. Denatured mRNA was size fractionated on a 1% agarose
                     gel. First strand cDNA synthesis was primed with an
                     oligo-dT primer containing a Not I site.  Double stranded
                     cDNA was size selected according to mRNA size fraction,
                     ligated with EcoR I adaptor, digested with Not I, and then
                     cloned directionally into pYX-Asc vector. The library tag
                     sequence located between the Not I site and the polyA tail
                     , is GTGCGTGGAA. This library was created for the
                     University of Iowa Mouse Brain Molecular Anatomy Project
                     (BMAP): 'Gene Discovery in the Developing Mouse Nervous
                     System', supported by National Instutues of Mental Health
                     (NIMH), Hemin Chin, Ph.D., program coordinator."
BASE COUNT      218 a    151 c    202 g    108 t
ORIGIN

Alignment Scores:
Pred. No.:               8.14e-51      Length:        679
Score:                   744.00        Matches:       164
Percent Similarity:      76.65%        Conservative:  10
Best Local Similarity:   72.25%        Mismatches:    47
Query Match:             9.22%         Indels:        7
DB:                      14            Gaps:          2

US-09-902-432-4 (1-1596) x BQ443215 (1-679)

Qy    1212 GluProThrGlnGlnGluLeuThrAlaAlaAlaValProValLeuAlaLysThrGluVal 1231
            ||||||||   |||   |||   |||   ||||||||:::||||||||   ||||||
Db       2 GAACCCACA---CAAGCGCTGGCTGCCGGGGCAGTGCCTATTCTGGCAAAGGCTGAGGTG 58

Qy    1232 GlyGlnGluGlyGluValAspTrpLeuAspGlyGluLysValLysGlu------------ 1247
            ||||||||||||||               ||||||||||||||||||:::
Db      59 GGTCAAGAGGGTGAGGCTGGCCAGTTTGATGGAGAAAAAGTCAAAGACGGACAGTGTGTT 118

Qy    1248 ---GluGlnGluValPheValHisSerGlyProAsnSerGlnLysAlaAlaAspValThr 1266
              |||  ||||||  ||||||:::||||||||||||||||   ||||||:::|||
Db     119 AAAGAACTGGAGGTGCCTGTGCACACTGGACCCAACAGTCAAAAGACTGCTGACTTGACA 178

Qy    1267 TyrAspSerGluValMetGlyValAlaGlyCysGlnGluLysGluSerThrGluValGln 1286
               ||||||||||||||   ||||||   |||||||||   ||||||   |||   |||
Db     179 CGTGACAGTGAAGTAATGGAAGTGGCCAGATGTCAGGAAACTGAGAGTAATGAAGAACAG 238

Qy    1287 SerLeuSerLeuGluGluGlyGluMetGluThrAspValGluLysGluLysArgGluThr 1306
           |||:::|||   |||:::   ||||||   |||||||||||||||||:::   ||||||
Db     239 AGTATTAGCCCGGAGAAAAGAGAGATGGGAACCGACGTTGAAAAGGAGGAAACAGAGACC 298

Qy    1307 LysProGluGlnValSerGluGluGlyGluGlnGluThrAlaAlaProGluHisGluGly 1326
           |||   ||||||   |||||||||   |||||||||||||||||||||||||||||||
Db     299 AAGACAGAGCAAGCCAGTGAAGAACATGAGCAGGAAACAGCTGCTCCTGAGCATGAAGGA 358

Qy    1327 ThrTyrGlyLysProValLeuThrLeuAspMetProSerSerGluArgGlyLysAlaLeu 1346
           |||:::   ||||||||||||||||   |||||||||   ||||||||||||||||||
Db     359 ACCCACCCTAAGCCAGTCCTGACAGCTGACATGCCTCACTCAGAGAGGGGAAAGGCACTG 418

Qy    1347 GlySerLeuGlyGlySerProSerLeuProAspGlnAspLysAlaGlyCysIleGluVal 1366
           |||||||||   ||||||||||||||||||||||||||||||   ||||||||||||
Db     419 GGCAGCCTTGAAGGAAGCCCTTCTCTCCCAGACCAAGACAAAGCAGATTGCATAGAGGTT 478

Qy    1367 GlnValGlnSerLeuAspThrThrValThrGlnThrAlaGluAlaValGluLysValIle 1386
           ||||||||||||   ||||||   ||||||||||||   |||||||||:::||||||
Db     479 CAAGTTCAAAGCTCAGACACACCAGTCACTCAAACAACCGAAGCTGTGAAAAAGGTCGAA 538

Qy    1387 GluThrValValIleSerGluThrGlyGluSerProGluCysValGlyAlaHisLeuLeu 1406
           |||||||||   ||||||   ||||||   ||||||   ||||||         |||
Db     539 GAAACTGTGGCAACTTCAGAGATGGATGAAAGTTTGGAGTGTGCAGGTGCGCAATCATTA 598

Qy    1407 ProAlaGluLysSerSerAlaThrGlyGlyHisTrpThrLeuGlnHisAlaGluAspThr 1426
           ||||||||||||   |||   |||||||||:::   ||||||||||||   |||   |||
Db     599 CCAGCTGAGAAGCTCTCCGAAACCGGTGGCTACGGGACTCTTCAGCATGGAGAGGA-ACC 657

Qy    1427 ValProLeuGlyProGluSer 1433
           ||||||   ||||||||||||
Db     658 GTGCCCCAGGGGCCTGAGTCT 678

RESULT 15
BF400579/c
LOCUS       BF400579               573 bp    mRNA    linear   EST 28-NOV-2000
DEFINITION  UI-R-CA0-bhd-a-09-0-UI.s1 UI-R-CA0 Rattus norvegicus cDNA clone
            UI-R-CA0-bhd-a-09-0-UI 3', mRNA sequence.
ACCESSION   BF400579
VERSION     BF400579.1  GI:11388554
KEYWORDS    EST.
SOURCE      Norway rat.
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 573)
  AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
  TITLE     Normalization and subtraction: two approaches to facilitate gene
            discovery
  JOURNAL   Genome Res. 6 (9), 791-806 (1996)
  MEDLINE   97044477
COMMENT     Contact: Soares, MB
            Program for Rat Gene Discovery and Mapping
            University of Iowa
            451 Eckstein Medical Research Building Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: msoares@blue.weeg.uiowa.edu
            The sequence contained an oligo-dT track that was pres
            oligonucleotide that was used to prime the synthesis o
            strand cDNA and therefore this may represent a bonafide
            tail. cDNA Library Preparation: M.B. Soares Lab Clone d
            clones will be available through Research Genetics (www
            Seq primer: M13 Forward
```

```
            POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1..573
                     /organism="Rattus norvegicus"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CA0-bhd-a-09-0-UI"
                     /clone_lib="UI-R-CA0"
                     /lab_host="DH10B (Life Technologies)"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-CA0
                     library is a subtracted library derived from the following
                     tissues: thalamus, cerebellum, hypothalamus, medulla, pons
                     , midbrain,  cerebral cortex, corpus striatum, testis, and
                     hippocampus.   For a detailed description of the library
                     from which this clone was  derived, please visit our web
                     site at ratest.eng.uiowa.edu.   The subtraction has been
                     previously described in (Bonaldo,  Lennon and Soares,
                     Genome Research 6:791-806, 1996)
                     TAG_SEQ=None found"
BASE COUNT        91 a    148 c    124 g    210 t
ORIGIN

Alignment Scores:
Pred. No.:               2.25e-48        Length:          573
Score:                   713.00          Matches:         143
Percent Similarity:      84.54%          Conservative:    21
Best Local Similarity:   73.71%          Mismatches:      24
Query Match:             8.83%           Indels:          6
DB:                      12              Gaps:            3

US-09-902-432-4 (1-1596) x BF400579 (1-573)

Qy     125 SerGluIleIleGluGlnIleProAlaSerGluAsnAsnValGluGluMetValGlnPro 144
           :::||||||||||||||||||||||:::||||||:::|||:::||||||:::     ||||||
Db     564 GCCGAAATAATCGAACAGATTCCTTCTTCAGAAAGCAATTTAGAAGAGCTAACACAACCC 505

Qy     145 AlaGluSerGlnAlaAsnAspValGlyPheLysLysValPheLysPheValGlyPheLys 164
                |||||||||||||||||:::|||||||||||||||||||||||||||||||||||||
Db     504 ACTGAGTCCCAGGCTAATGATATTGGATTTAAGAAGGTGTTTAAGTTTGTTGGCTTTAAA 445

Qy     165 PheThrValLysLysAspLysAsnGluLysSerAspThrValGlnLeuLeuThrValLys 184
           ||||||||||||||||||||     ||||||   ||||||||||||||||||||||||||
Db     444 TTCACTGTGAAAAAGGATAAGACAGAGAAGCCTGACACTGTCCAGCTACTCACTGTGAAG 385

Qy     185 LysAspGluGlyGluGlyAlaGluAlaSerValGlyAlaGlyAspHisGlnGluProSer 204
           |||||||||||||||||||||               ||||||||||||||||:::||||||
Db     384 AAAGATGAAGGGGAGGGAGCA---------GCAGGGGCTGGCGACCACCAGGACCCCAGC 334

Qy     205 ValGluThrAlaValGlyGluSerAlaSerLysGluSerGluLeuLysGlnSerThrGlu 224
           :::               ||||||:::|||||||||||||||||   |||||||||||||||
Db     333 CTT------GGGGCTGGAGAAGCAGCATCCAAAGAAAGCGAACCCAAACAATCTACAGAG 280

Qy     225 LysGlnGluGlyThrLeuLysGlnGluGlnSerSerThrGluIleProLeuGlnAlaGlu 244
           |||   |||   |||||||||:::|||||||||      ||||||         ||||||
Db     279 AAACCCGAAGAGACCCTGAAGCGTGAGCAAAGCCACGCAGAAATTTCTCCCCCAGCCGAA 220

Qy     245 SerAspGlnAlaAlaGluGluGluAlaLysAspGluGlyGluGluLysGlnGluLysGlu 264
           |||   ||||||      ||||||   |||:::|||||||||||||||||||||||||||
Db     219 TCTGGCCAAGCAGTG---GAGGAATGCAAAGAGGAAGGAGAAGAGAAACAAGAAAAAGAA 163

Qy     265 ProThrLysSerProGluSerProSerSerProValAsnSerGluThrThrSerSerPhe 284
           |||:::||||||   |||||||||:::|||||||||   |||||||||   |||:::|||
Db     162 CCTAGCAAGTCTGCAGAATCTCCGACTAGTCCCGTGACCAGTGAAACAGGATCAACCTTC 103

Qy     285 LysLysPhePheThrHisGlyTrpAlaGlyTrpArgLysLysThrSerPheLysLysSer 304
           |||||||||||||||   |||||||||||||||||||||||||||||||||||:::|||
Db     102 AAAAAATTCTTCACTCAAGGTTGGGCCGGCTGGCGCAAAAAGACCAGTTTCAGGAAGCCG 43

Qy     305 LysGluAspAspLeuGluThrAlaGluLysArgLysGluGln 318
           |||||||||:::::::|||   :::||||||:::|||::::::
Db      42 AAGGAGGATGAAGTGGAAGCTTCAGAGAAAAAAAAAAAAAAA 1
```

```
Search completed: December 13, 2002, 04:10:45
Job time : 4157 secs
```